

THE EDUCATION REFORM OF TAP AND VALUE-ADDED ASSESSMENT:  
TEACHER MERIT PAY THAT REINVIGORATES STANDARDIZED TESTING  
AND DETRACTS FROM 21ST CENTURY LEARNING SKILLS

A Thesis

presented to

the Faculty of California Polytechnic State University,

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts in History

by

Shawn Scharer Greenelsh

Shawn Scharer Greenelsh

ALL RIGHTS RESERVED

June 2011

©2011

## COMMITTEE MEMBERSHIP

TITLE: The Education Reform of TAP and Value-Added Assessment: Teacher Merit Pay that Reinvigorates Standardized Testing and Detracts from 21st Century Learning Skills

AUTHOR: Shawn Scharer Greenelsh

DATE SUBMITTED: June 2011

COMMITTEE CHAIR: Dr. Tom R. Trice, Associate Professor

COMMITTEE MEMBER: Dr. Joel Orth, Assistant Professor / Social Science Teaching Credential Advisor / History Minor Advisor

COMMITTEE MEMBER: Dr. Roberta J. Herter, Assistant Professor

## ABSTRACT

TITLE: The Education Reform of TAP and Value-Added Assessment: Teacher Merit Pay that Reinvigorates Standardized Testing and Detracts from 21st Century Learning Skills

Shawn S. Greenelsh

In the last two decades, ignoring the bulk of educational research findings, policymakers shaped educational policy into a standardized testing movement that now dominates education. Now, to comply with *No Child Left Behind*, teachers and administrators shape curriculum in a way that maximizes student achievement measured by these tests. Recently, business and educational leaders initiated a reform movement to broaden curriculum, narrowed by this inadequate standardized testing movement, so that necessary 21st century learning skills can be practiced through project-based learning. The Federal Government's enforcement of power over education created the climate that defined the current educational policy that gave birth to the standardized testing movement. In this climate, this reform to boost 21st century learning skills does not gain practical traction that results in changed policy, because it is impossible for standardized testing to assess most of these skills and this type of learning due to the limitations of bubbled-multiple choice questions.

Instead of shaping policy to foster these 21st century learning skills, policymakers push another reform, through TAP (The System for Teacher and Student Advancement) and Value-Added Assessment. This reform attempts to improve instruction through

teacher merit pay--a device that has failed many times in educational reform history.

Unfortunately, most TAP systems use standardized tests as the only student achievement measurement, so almost all student achievement gains involving 21st century learning skills and project-based learning are not officially measured. Efforts to use portfolios and authentic assessment, the measurement tools that should be used to measure these higher level skills, are not supported by policymakers, because the lack of standardization requires more trust in the assessment ability of local school districts and communities.

Consequently, a massive disconnect exists where standardized testing is being reinvigorated instead of de-emphasized, and this comes with the potential price of many teachers and administrators not embracing 21st century learning skills and project-based learning as much as they could if they were not bound by standardized test results.

Ultimately, these two reforms that contradict each other involve larger issues of jurisdictional power over education at federal, state, and local levels, and ideological challenges to teacher job security and teacher representation.

## TABLE OF CONTENTS

### CHAPTER

I.	INTRODUCTION . . . . .	1
	Purpose and Methodology . . . . .	6
	Survey of Initial Sources . . . . .	12
II.	A BRIEF HISTORY OF STANDARDIZED TESTING . . . . .	15
III.	THE INADEQUACY OF STANDARDIZED TESTS . . . . .	31
IV.	THE PUSHING AND RESISTING OF TEACHER MERIT PAY . . . .	62
V.	THE TAP SYSTEM, THE DIRECTION OF EDUCATION IN SYCAMORE RIVER AND THE NATION, AND RELATIVE TEACHER PERCEPTIONS . . . . .	100
VI.	BIBLIOGRAPHY . . . . .	135
VII.	APPENDICES	
	A. Survey Responses . . . . .	145
	B. Survey Comments . . . . .	147
	C. Interview Details and Format . . . . .	151
	D. A Brief Description of TAP . . . . .	152

## Chapter I Introduction

We live in a new century, in a new world. In the last twenty years, technology connected the world creating a globalized village where ideas flow among all people in milliseconds. In the last decade, simple hand-held devices provide PC-like power and connect us to the internet. Within seconds a user can look up information about any topic, define any term, explain concepts, or provide historical background. The latest demographic trends, the way superconductors work, and wholesale suppliers of Mount Kilimanjaro coffee beans are all examples of the most up to date information provided by this hand-held technology, which in 2011 holds more information than the entire human race possessed in 1911. No longer do we live in a world where the access to knowledge inhibits people. What we do with the information defines what counts. Taking knowledge and building something new, conceptually or practically, requires critical thinking, synthesis, and collaboration. Education today, more than any other day, must fulfill a new role in teaching students how to best use this readily available information.

I ponder my fifteen year teaching career and realize that, much like a doctor practices medicine and a lawyer practices law, my career humbly unfolds on a path of improvement that will never end with perfection, and where new lessons appear daily. Thus far, many important lessons form my teaching foundation. Students are extremely dynamic creatures, and a student's learning capacity daily depends on the teacher, classroom, and school, but also on the home-life of the student, something that is out of my control. Providing students power to problem-solve affects their life much more than giving them the answer, because the process of finding the answer turns them into self-directed learners. Teachers will not be in front of them forever, and there is so much to

learn, they will have to grow into lifelong learners to prosper in this new world.

Listening to students guides me to adjust my teaching to meet their needs. Most importantly, I am not alone. Other teachers trod a similar path filled with similar lessons.

The power of project-based learning repeatedly demonstrates to me a type of teaching that brings to fruition other major lessons learned over the years. This type of learning opens an avenue that places students in situations or scenarios where they practice critical thinking, self-directed learning, collaboration, and communication so they are prepared to navigate through today's demanding world. Project-based learning is learning that is driven by an open-ended question that formulates a problem that requires the students to find a solution.<sup>1</sup> The process of finding a solution requires the students to research information and critically think through the possible solutions to the problem. Usually, students work in teams, and then communicate their findings to the

---

<sup>1</sup> Here is a list of contributors to my ideas behind project-based learning:

The Buck Institute for Education is a great resource for anyone interested in more information about project-based learning. As a non-profit organization since 1987, they offer, free of charge, a library of projects, developed over two decades, useful for teachers just starting project-based learning. I received valuable professional development training from two Buck Institute trainers contracted by Sycamore River. The website for the Buck Institute for Education is: <http://www.bie.org/> (accessed January 20, 2011).

Deborah Meier, *The Power of Their Ideas: Lessons for America from a Small School in Harlem* (Boston: Beacon Press, 2002).

Mary Ellen Freeley, and Richard Hanzelka, "Getting Away from Seat Time: A New Hampshire initiative encourages schools to move toward competency-based learning," *Educational Leadership* 67, no. 3 (November, 2009): 63-67.

Partnership for 21st Century Learning Skills. <http://www.p21.org/> (accessed November 23, 2010).

Bob Pearlman, "21st Century Learning in Schools--A Case Study of New Technology High School in Napa, CA," <http://www.bobpearlman.org/Articles/21stCenturyLearning.htm> (accessed May 12, 2011).

Bernie Trilling, and Charles Fadel, *21st Century Skills: Learning for Life In Our Times* (San Francisco: Jossey-Bass, 2009).

Chandra J. Foote, Paul J. Vermette, and Catherine F. Battaglia, *Constructivist Strategies: Meeting Standards and Engaging Adolescent Minds* (Larchmont: Eye on Education, 2001).

Pearl G. Solomon, *The Curriculum Bridge: From Standards to Actual Classroom Practice* (Thousand Oaks: Corwin Press, 2009).

David H. Jonassen, and Susan M. Land, *Theoretical Foundations of Learning Environments* (Mahwah: Lawrence Erlbaum Associates, Publishers, 2000).

Daniel M. Savage, *John Dewey's Liberalism: Individual, Community, and Self-Development* (Carbondale: Southern Illinois University Press, 2002).



teacher and the rest of the class. The findings are open to scrutiny by the teacher and the rest of the class in a Socratic format. The solution becomes, in a way, a universal lesson applicable to different locations and time periods.

A fore-bearer to this type of project-based learning is a teacher and author, Deborah Meier. She documented her experience in her book, *The Power of Their Ideas: Lessons for America from a Small School in Harlem*, to share the details of the creation of an alternative approach to the standards-based learning method. She and several other teachers created a primary school in 1974, followed later by a secondary school, using a project-based learning model. These two schools are *Central Park East*, in New York. The first high school graduating class was 1991, and in the last two decades both these schools have become models for how “inner-city” schools can succeed despite a myriad of obstacles. These schools follow the philosophies of John Dewey and Jean Piaget, by striving to combine democracy with education. Meier explained society’s mixed feelings about this approach, “Small, democratically run schools are both quintessentially American and hard for Americans to swallow. They appeal to our spirit of independence, but not to our impatient desire for guaranteed fixes and standardized products.”<sup>2</sup> In this democratic setting, the students use inquiry methods to tackle cross curricular tasks with the freedom to explore various solutions. This approach mirrors the way scientists search for a solution in a lab, or lawyers investigate evidence for a trial, or historians research primary sources<sup>3</sup>. The inquiry method is much the way adults in a democracy solve problems.

---

<sup>2</sup> Deborah Meier, *The Power of Their Ideas: Lessons for America from a Small School in Harlem* (Boston: Beacon Press, 2002), 37.

<sup>3</sup> Meier, 50.

Two methods help guide the students in their search to solve the problems required to complete the tasks. First, the students use what is called *Habits of Mind*. These five habits are: *evidence*-how do you know that; *viewpoint*-who said it and why; *cause and effect*; what led to it, what else happened; *hypothesizing*-what if, or supposing that; and *habit*-who cares. Meier explained that the last one is the most important. She explained, “Knowing and learning take on importance only when we are convinced it matters, it makes a difference...It matters because it will help us get ahead, get into a good college, hold a well-paying job...It will also help save the world.”<sup>4</sup> The graduates of Central Park East use this disciplined approach to the inquiry method to accomplish the tasks necessary for college and career.

Another method to solve problems and complete projects is the *Habits of Work*. According to Meier, the *Habits of Mind* don’t mean very much if the quality of product and the process of working with others is not a concern. The *Habits of Work* are effective because the students practice at “the acceptance of increasing levels of responsibility, the increasing capacity to communicate appropriately to others, a willingness to take a stand as well as a willingness to change one’s mind, and being someone who can be counted on to meet deadlines as well as keep one’s word.”<sup>5</sup> The students can step back and evaluate a job well done from a project that took significant time and effort. The students start to see real results and the cycle of success becomes intrinsically motivated.

The graduation requirements from Central Park East Secondary School (CPESS) require the students to demonstrate a mastery over the *Habits of Mind* and the *Habits of Work*. Meier explained the difference between standard high schools and the more alternative approach, “...portfolio-based graduation requires our students to prepare tangible demonstrations of their knowledge and competence rather than accumulating

---

<sup>4</sup> Meier, 41.

<sup>5</sup> Meier, 49.

‘seat-time’ (credits) or grades on multiple choice tests.”<sup>6</sup> History, ethics, science, math, and media are among the fourteen different portfolio areas that define the project-based learning process for seniors to graduate.<sup>7</sup> The seniors work on campus independently and in collaboration with other students, with teachers’ help at school, and on their own, off campus. The portfolios are regularly reviewed “...by panels of experts consisting of college faculty, high school colleagues, parents, community leaders, discipline experts, and educational policy-makers and officials.”<sup>8</sup> Instead of standardized tests that are run through a computer, there are real people connected to the community and the school that are seeing and admiring what the students creatively and energetically produced.

Project-based learning can occur in a school-wide effort like Central Park East or more randomly in a class-by-class approach. In a single classroom, for example, when studying the Progressive Era in an Eleventh Grade U.S. History class, a project-based learning question or task could be to develop and implement a way to make a positive difference with a social problem that was around during the Progressive Era and is still around today. Women’s struggle for equality could be a topic students focus on, with a current issue being domestic violence. The project students choose could be to create a public awareness campaign about domestic violence and implement a food and clothing drive for a local women’s shelter. The details of the Progressive Era can be accessed by the click of a few buttons, but the process of synthesizing the information, negotiating and communicating with other students and community members, and critically evaluating options, builds skills in the students that will guide and encourage them to continue to be self-directed learners and critical thinkers for the rest of their lives. To measure student achievement growth from this project requires much more than anything

---

<sup>6</sup> Meier, 30.

<sup>7</sup> Meier, 42.

<sup>8</sup> Meier, 57.

that can be bubbled on a standardized test. It takes comparing pre and post presentations or comparing previous artifacts in a portfolio. This is known as authentic assessment. It is much more difficult to evaluate beyond a local level compared to standardized testing, and therefore state and national education policy pushes it aside in favor of standardized testing. These jurisdictional issues lead to bad policy.

### Purpose and Methodology

In America today, there exists a disconnect between educational research and education policy. Educational research clearly indicates the advantages of authentic assessment in measuring student and teacher performance, and yet education policy is becoming more focused on standardized testing. Authentic assessment and standardized testing are historically grounded on two different approaches to learning. For various reasons, standardized testing gained more ground through the years, and now in the last two decades, this divide, between educational research, which calls for authentic assessment, and education policy, which drives standardized testing across the United States, has grown so large that it leaves you baffled, scratching your head, wondering what are the policymakers thinking. If the research findings are so conclusive, then why do policies pushing standardized testing continue? Especially when considering that this disconnect could prevent education from reaching an important reformation, from an educational system dominated by standardized testing to an educational system designed to meet the current demand for critical thinking, self-directed learners. Surely the policymakers would listen to parents if they provided a critical mass of opposition, but how will the parents know about this disconnect if they only hear the rhetoric from

policymakers? Do teachers even realize the magnitude of the divide between the educational research findings and policy? This study sets out to explain the causes of this disconnect and then to explore, why at this point in history, the divide grows larger instead of closing.

Educational science and education policy developed two distinct discourses on best practices, which has resulted in the last two decades in a massive disconnect between research findings and policymaking, both nationally and locally. Researchers have known for decades that standardized testing is preventing educators from teaching more critical thinking, self-directed learning, and yet education policy continues to raise the stakes connected to standardized testing results. The history of this disconnect between the educational research and education policy shows that standardized testing and project-based learning results have been mutually opposed or inconsistent.

To understand this disconnect, this paper exposes research showing the inadequacy of standardized testing and the reasons why teacher merit pay connected to these test results will not work. Then, the paper will use one school district as an example of how standardized testing and teacher merit pay impact locally. For the purpose of this paper the school district name has been changed to Sycamore River Unified School District.<sup>9</sup> The disconnect between educational research and education policy is evident in Sycamore River through its use of two organizations, *The System for Teacher and Student Advancement* (TAP) and *Partnership for 21st Century Learning* (P21), that are incompatible with each other. These two organizations push educators in

---

<sup>9</sup> All names and identifying characteristics of local individuals and the school district in focus have been changed to protect the anonymity of the participants in the study. Words from the district website and from individual interviews are used, but the district website and the names of these individuals will not be footnoted for the same reasons of protected anonymity.

two different directions. TAP places more emphasis on standardized test results, and P21 places more emphasis on project-based learning. Consider the next two statements. The first statement is from the TAP brochure, and the second statement is from the Evergreen Freedom Foundation which endorses the Value-Added Assessment system that TAP uses to evaluate teachers. Both of these organizations fully support standardized testing. On the surface, both these statements appear very legitimate, but if the goal is student achievement measured only by standardized test results, then, according to educational research in the last two decades, the learning, even taught by the best teachers, will be riddled with limitations.

When it comes to student achievement—across all socioeconomic levels—talented teachers are the critical factor. That’s right. Within the course of the school day, nothing has more impact on student learning than the effectiveness of the teacher in the classroom. Decades of well-intentioned but piecemeal reforms failed to significantly raise student achievement because they also failed to make teacher excellence the cornerstone of school reform.<sup>10</sup>

Achievement tests seem to draw fire from all sides, but in one form or another, they must be administered. Parents, teachers, legislators, and businesses want a legitimate way to evaluate a school system’s effectiveness. Parents and teachers in particular want to know how much children really know in comparison to academic standards. These goals require instruments that measure the achievement level of individual children as well as schools in a given group. To be fair, the instruments must be “standardized”: tests must be uniformly administered, identical tests, or tests made equivalent, must be offered in different locations and years, and scores must be expressed in a standard way that allows for fair comparisons.<sup>11</sup>

Now consider two statements that originate from supporters of 21st century learning skills. The first statement comes from a book that the Superintendent of

---

<sup>10</sup> TAP--The System for Teacher and Student Advancement-brochure [http://www.tapsystem.org/pubs/tap\\_brochure.pdf](http://www.tapsystem.org/pubs/tap_brochure.pdf) (accessed June 22, 2011). TAP is now run through the National Institute for Excellence in Teaching.

<sup>11</sup> Evergreen Freedom Foundation. “Value Added Assessment.” *Evergreen Freedom Foundation, School Director’s Handbook* (n.d.) <http://www.myfreedomfoundation.com/pdfs/Value-Added.pdf> (accessed October 25, 2010).

Sycamore River, recommended to his teachers in August 2009, in a districtwide beginning of the school year presentation about 21st century learning skills. This book shows the problems with the instruction shaped around standardized testing, which is the same type of instruction that Sycamore River is pushing with the TAP system and Value-Added Assessment. The second statement comes from the Buck Institute for Education webpage that describes what project-based learning should look like:

In many ways, standards have been designed for the way we test. Standards have been limited to the types of knowledge best tested by the multiple-choice questions on the machine-scored tests so commonly used to measure student progress. [new paragraph] This has led teachers to focus on “coverage,” superficially rushing through a vast number of topics with their students, and to emphasize memorization and recall in preparation for the end-of-year, high-stakes standards-based tests that determine so much of a student’s future learning path.<sup>12</sup>

**Requires critical thinking, problem solving, collaboration, and various forms of communication.** Students need to do much more than remember information—they need to use higher-order thinking skills. They also have to learn to work as a team and contribute to a group effort. They must listen to others and make their own ideas clear when speaking, be able to read a variety of material, write or otherwise express themselves in various modes, and make effective presentations. These skills, competencies and habits of mind are often known as “21st Century Skills.”

**Allows some degree of student voice and choice.** Students learn to work independently and take responsibility when they are asked to make choices. The opportunity to make choices, and to express their learning in their own voice, also helps to increase students’ educational engagement. **Incorporates feedback and revision.** Students use peer critique to improve their work to create higher quality products. **Results in a publicly presented product or performance.** What you know is demonstrated by what you do, and what you do must be open to public scrutiny and critique.<sup>13</sup>

Many teachers and community members in the Sycamore River school district want to move in the direction of teaching more 21st century learning skills through

---

<sup>12</sup> Bernie Trilling, and Charles Fadel, *21st Century Skills: Learning for Life In Our Times* (San Francisco: Jossey-Bass, 2009), 126.

<sup>13</sup> Buck Institute for Education--information about what project-based learning looks like. [http://www.bie.org/about/what\\_is\\_pbl](http://www.bie.org/about/what_is_pbl) (accessed June 22, 2011). Bold print apart of original text.

project-based learning.<sup>14</sup> With the Superintendent's leadership, Sycamore River district selected a set of goals that I believe will prepare our students for today's competitive world. By using ideas from P21, Sycamore River developed a *21st Century Learning Skills Framework* that identifies a list of skills students need to be successful in our world today.<sup>15</sup> *Sycamore River Unified School District Vision 2020 Framework* sets a goal that states that students will be able to: communicate, problem solve, and critically think through self-directed learning, and through teamwork, collaboration, and cooperation with the use of technology, innovation, imagination, and creativity. Sycamore River's recognized mission is to engage, challenge, and inspire students through the power of learning.

Today, Sycamore River experiences the same disconnect that has plagued educators across the nation for decades. Months after this *Vision 2020 Framework* was published, Sycamore River received a Teacher Incentive Fund (TIF) grant from the federal Department of Education, in the amount of 7.2 million dollars. This grant stipulates that teachers at six school sites will be evaluated for performance pay using a new system, called TAP (The System for Teacher and Student Advancement). Within this new evaluation system, half of a teacher's evaluation will be determined by student achievement growth using a tool called Value-Added Assessment. The majority of TAP schools across the country (which comprises 200,000 students and 20,000 teachers) use

---

<sup>14</sup> From results of a district wide teacher and community survey asking: "What skills do our students need to be successful in the 21st Century?" There were approximately 250 people who were surveyed. The results were published June 2010 on the district website.

<sup>15</sup> Many of the ideas that Sycamore River uses regarding 21st Century Learning Skills come from the Partnership for 21st Century Learning Skills. <http://www.p21.org/> (accessed November 23, 2010). The ideas also come from a book by Bernie Trilling and Charles Fadel. *21st Century Skills: Learning for Life In Our Times*. San Francisco: Jossey-Bass, 2009.



scores from standardized tests to determine what student achievement growth has occurred.<sup>16</sup>

Just when I thought standardized testing would be de-emphasized in Sycamore River, to focus more on project-based learning and 21st century learning skills, the TIF grant came along, requiring more emphasis on standardized testing. Sycamore River is a test case for California. Until now, TAP has never been used in California.<sup>17</sup> The overarching question that only time will answer is: Will this teacher performance-pay system, that relies heavily on the scores from standardized testing, define the direction education takes in Sycamore River and other areas of the United States. Sycamore River is at a crossroads and I would like to provide a map to help inform its choice; like any good map, it needs to show where we are now, how we got there, and where the roads take us.

This disconnect between educational research and educational policy has been present for decades. The developing pedagogical debate in education can be tracked using various sources written by educational experts and educational policy consultants, which are educational journals, private think-tank papers, and books written by experts in the field, who are usually professors of education, history, sociology, business, and economics. Some examples used are: Brookings Institute, Educational Testing Service, National Education Association, National Council on Measurement in Education,

---

<sup>16</sup> Information regarding the TIF grant and the TAP evaluation system, specific to Sycamore River, can be obtained from Sycamore River Unified School District's home page on its website [district website omitted for anonymity reasons] A TAP representative to Sycamore River, November 12, 2010, explained to a group of teachers at a middle school, that a majority of TAP sites use standardized testing, because it is already in place and it is easy to use.

<sup>17</sup> John Fensterwald, "Experiments in evaluating teachers: Districts and charters breaking new ground," Thoughts On Public Education, Simi Valley Education Foundation (June 1, 2011), <http://toped.svefoundation.org/2011/06/01/experiments-in-evaluating-teachers/> (accessed June 14, 2011). This website contains a very informative blog about how readers weigh-in on this issue.

American Council on Education, Education in American History, Evergreen Freedom Foundation, National Society for the Study of Education, Performance and Assessment of California Teachers, California Standards for the Teaching Profession, California State Standards and Instructional Practices, Teacher Advancement Program Leadership Handbook, Sycamore River Unified School District's Teacher Incentive Fund (TIF) Grant Application, Center for Educator Compensation Reform, The Journal of Economic Perspectives, and the Yearbook of the National Society for the Study of Education.

A variety of other sources show this disconnect within Sycamore River. Some of these sources will be: district communication related to the TIF grant and TAP evaluation system in Sycamore River from the teacher's union and the district leadership, a Power-Point presentation from a TAP representative, communication regarding P21, observations from the project-based learning class that I implemented beginning in the 2010/2011 school year, observations from my role on the Meet and Confer Committee--formed to represent teachers regarding the negotiations involved in adopting the TAP evaluation system in Sycamore River, observations from a visit to a New Tech High in Austin, Texas, survey responses from 37% of Sycamore River teachers, and 25 interviews, comprising over four hours of videotaped oral history of teachers, counselors, and administrators regarding how standardized testing has affected Sycamore River over the last twenty years.

### Survey of Initial Sources

When I gathered sources for this study, I didn't know what the research said, exactly. I grabbed whatever sources were available regarding the general topics of

standardized testing, teacher merit pay, project-based learning, and authentic assessment.

Before I began my study, I had obtained approximately 90 sources. A survey of the sources provided a quantification of the research findings that were generally accessible through the Cal Poly, San Luis Obispo Library, Interlibrary Loan Services, and articles found through Google Scholar. For an abbreviated purpose, in parenthesis, the author's last name and the first two words of the title are provided to be able to find the source in my bibliography:

---

I. Regarding the history of accuracy and authenticity of standardized testing--thirty sources:

Twenty-two provided an argument against standardized testing:

(Rothstein--*Grading Education*; Ravitch--*The Revisionists*; Ravitch--*Brookings Papers*; Spring--*The Sorting*; Shavelson--*A Brief*; Dorn--*Accountability Frankenstein*; Solomon--*The Curriculum*; Ladd--*Holding Schools*; Ravitch--*The Death*; Ryan--*The Future*; Baker--*Problems with*; Rothstein--*What Do*; Kohn--*Fighting the*; Sacks--*Standardized Testing*; Kane--*The Promise*; Olson--*Tennessee Reconsiders*; Doran--*Challenges of*; Loveless--*Test Based*; Nichols--*Collateral Damage*; Brookhart--*The Many*; Bracey--*Big Tests*; Braun--*Getting Value*)

Five provided for and against arguments regarding standardized testing:

(Brennan--*Educational Measurement*; Ravitch--*Debating the*; Cohen--*Education and*; Barton--*A Policy...Facing*; Evans--*Taking Sides*)

Three supported standardized testing, but an author of one of the sources, Diane Ravitch, changed her mind six years after the publication of that book to argue against standardized testing: (Ravitch--*Left Back*; Phelps--*Why Testing*; Peterson--*Choice and*)

---

II. Regarding connecting teacher merit pay to standardized test results--thirty-eight sources:

Sixteen argued against connecting teacher merit pay to standardized test results:

(Urban--*Old Wine*; Wilms--*The Illusion*; Baker--*Problems with*; Barton--*A Policy...Order in*; Cresap--*Teacher Incentives*; Parker--*Career Ladder*; Haertel--*Uses and*; Springer--*Teacher Pay*; Clotfelter--*Do School*; Brandt--*Incentive Pay*; Fisk--*When Schools*; Gipps--*Accountability Testing*; Larabee--*Can Teachers*; Herndon--*Merit Pay*; Wood--*Managerial Experience*)

Eleven explained ways to encourage increased teacher performance without using standardized test results as a high-stakes factor:

(Bacharach--*Evaluating Teachers*; Darling-Hammond--*A License*; Ladson--*No Teacher*; Moulthrop--*Teachers Have*; National Center--*Tough Choices*; Odden--*Paying Teachers*; PACT--*Performance Assessment*; Rumery--*Measuring Teacher*; Wiley--*Denver Pro*; Gratz--*The Problem*; Springer--*Performance Incentives*)

Three showed why merit pay fails:

(Dockery--*The Teacher*; National Education--*SEARCH*; Van Loozen--*Some Points*)

Seven argued for a high-stakes connection of teacher merit pay to standardized test results:

(Bacharach--*Paying For*; Evergreen--*Value Added*; Solomon--*The Case*; Solomon--*The Pros*; Wiley--*A Practitioner's*; Wynn--*American Education*; Sanders--*Comparisons Among*)

One is a general reference for four models of alternative teacher compensation systems in use today: (Natale--*Retaining and*)

---

III. Regarding the topics of 21st century learning skills, project-based learning, and authentic assessments--twenty-three sources:

Five explained the details and importance of 21st century learning skills:

(*Partnership for 21st Century Learning Skills*; Sycamore River Unified School District; Barton--*A Policy--What Jobs Require*; Winger--*Grading What*; Trilling--*21st Century*)

Nine provided the history and details behind project-based learning:

(Tomlinson--*Edward Lee*; Steffy--*Curriculum and*; Senechal--*The Most*; Schwartz--*Constructivism in*; Savage--*John Dewey's*; Dwyer--*The Future*; Darling-Hammond--*Authentic Assessment*; Buck Institute for Education; Meier--*The Power*)

Nine explained why authentic assessment needs to grow in importance:

(Wilson--*Towards Coherence*; Tucker--*The Next*; Linn--*Measurement and*; Kohn--*The Case*; Johnson--*Merit, Money*; Holland--*Assessing the*; Freeley--*Getting Away*; Baker--*Understanding Educational*; Belanoff--*Portfolios: Process*)

---

## Chapter II

### A Brief History of Standardized Testing

Before the Information Age, the capacity to store information in one's memory was highly valued. Many standards today represent a carry-over from this time. Standardized tests today still reflect this value for learning information that has no real-life application; it is almost a mentality of valuing knowledge-for-knowledge-sake. The current standards-based learning, measured by standardized tests, is defined by standards that require time and energy memorizing information that could merely be accessed using technology. These are standards that lead to knowledge retrieval rather than critical thinking. Here are some examples of released questions from the California STAR test for Eleventh Grade Social Studies (I included questions related to the Progressive Era to help compare to the project-based learning example previously stated):

Which of the following was an effect of the publication of Upton Sinclair's *The Jungle* (1906)?

- A It aided the growth of federal social services.
- B It contributed to the development of settlement houses.
- C It influenced the passage of the Meat Inspection Act.
- D It led to the development of child labor laws.

The muckraking journalists associated with the Progressive Era were known *primarily* for their

- A willingness to expose the corruption of U.S. society.
- B articles supporting the economic benefits of laissez-faire economics.
- C use of the media to advocate the passage of the Equal Rights Amendment.
- D support for the formation of U.S. military alliances with European countries.<sup>18</sup>

Learning about Upton Sinclair and the Muckrakers can be important if the thematic lesson is that people in the past dedicated hard work and risk to stop corruption and abuse. The same theme could be used to teach 21st century learning skills with project-

---

<sup>18</sup> California Department of Education, "California Standards Test-Released Test Questions," pages 13-14 (March, 2009). <http://www.cde.ca.gov/ta/tg/sr/documents/cstrtqhssmar18.pdf> (accessed November 23, 2010).

based learning where the students are given the opportunity to become a modern-day Upton Sinclair or Muckraker by creating public awareness about corruption or abuse around us today. If the goal of standardized tests is to evaluate student achievement growth, then the question becomes: what is the student achievement that is being measured? If the goal is to store details about the contributions of Upton Sinclair or the Muckrakers, then the sample California standardized testing questions (stated above) are valid and authentic for measuring student achievement. However, if the goal is to measure whether students have become self-directed learners, able to use technology to access knowledge to create a solution, then those sample questions are *not* valid and authentic.

Many of the standards today are valuable, and do lead to teaching real-life application skills that align well with skills needed in the 21st century. There are also valuable historical lessons for how humans should treat each other and the environment around us. Real-life value can be found among many of the hundreds of standards across the subject areas. The problem lies in the fact that many of these valuable standards cannot be tested using standardized tests, because the assessment format does not fit multiple choice questions with bubbled-in responses.

A push-back against the barrage of standards has occurred within the last several years. The official statement from administrators used to be that teachers should attempt to cover nearly all the standards. It is estimated that in some cases “it would take as many as twenty-two years of schooling to adequately teach all the content identified in a set of elementary school standards documents!”<sup>19</sup> Now, the idea is to prioritize the

---

<sup>19</sup> Bernie Trilling, and Charles Fadel, *21st Century Skills: Learning for Life In Our Times* (San Francisco: Jossey-Bass, 2009), 126.

standards. These are known as “essential standards” or “power standards.”

Administrators now recognize that teachers must prioritize given the limited classroom-time in a year. The question becomes, how does one define “essential.” Some administrators would define it as the standards that will make the biggest difference on standardized tests, because that is where the stakes have been set by education policy at the state and national level, and that is where the administrators’ job performance is measured. However, there are valuable standards that prepare students to possess skills for success in life, but those standards will not be tested, and therefore risk not being considered “essential.” For example, the *California Standard 2.6, Grades Eleven and Twelve, English Language Arts Writing Applications Standards*, states that students should be able to:

Deliver multimedia presentations:

- a. Combine text, images, and sound and draw information from many sources (e.g., television broadcasts, videos, films, newspapers, magazines, CD-ROMs, the Internet, electronic media-generated images).
- b. Select an appropriate medium for each element of the presentation.
- c. Use the selected media skillfully, editing appropriately and monitoring for quality.
- d. Test the audience’s response and revise the presentation accordingly.<sup>20</sup>

Obviously, a student who can deliver a multimedia presentation is equipped with many skills needed in the 21st century, but unfortunately the student achievement growth that occurred within this multimedia presentation will never be reflected on any standardized test scores, because this type of learning cannot be measured with a multiple choice question.

---

<sup>20</sup> California State Board of Education-English-Language Arts Content Standards for California Public Schools-Kindergarten Through Grade Twelve (December, 1997).  
<http://www.cde.ca.gov/be/st/ss/documents/elacontentstnds.pdf> (accessed November 24, 2010)

To understand standardized testing today, it is helpful to survey its history. According to the research findings from Richard Rothstein, Rebecca Jacobsen, and Tamara Wilder, in their book, *Grading Education: Getting Accountability Right*, when studying the history of American public education, a pattern of expected learning outcomes emerges. From the early Republic, to today, an expectation of student learning outcomes delineates into eight general categories: basic academic knowledge and skills, critical thinking and problem solving, appreciation of the arts and literature, preparation for skilled employment, social skills and work ethic, citizenship and community responsibility, physical health, and emotional health.<sup>21</sup> The writings of early leaders like Benjamin Franklin, George Washington, and Thomas Jefferson are consistent with these stated outcomes. In the 1830s, Horace Mann, secretary of the newly created Massachusetts Board of Education, began a campaign to spread public education to more students. He also expressed a notion of desired outcomes for students which we might recognize today as a “standard.” The earliest American education “standards” came down to civic, moral, and academic knowledge and skills. The assessment of these “standards” were left up to local educators to design and implement.<sup>22</sup>

With the expansion of public education in the 1890s, came the first effort to standardize the learning in schools across the nation. Two different motivations led the effort to create standards. One pushed for more rigorous academic standards to ensure high school students were prepared for college, and the other pushed to apply the basic academic knowledge learned in primary school to learn vocational and civic skills.<sup>23</sup>

---

<sup>21</sup> Richard Rothstein, Rebecca Jacobsen, and Tamara Wilder, *Grading Education: Getting Accountability Right* (New York: Teachers College Press, 2008), 14.

<sup>22</sup> Rothstein, Jacobsen, and Wilder, 14-19.

<sup>23</sup> Rothstein, Jacobsen, and Wilder, 19-22.



Early in its history, American public schools struggled with focusing on one specific purpose or mission, which meant that there was no commonly identified outcomes to assess in a standard fashion. These mixed messages continued to pull educators in different directions through the twentieth century and to the present.

The supporters of more rigorous academic standards worked to implement standardized testing in the United States early in the twentieth century. Some of these tests eventually were used to determine if students reached an adequate level of learning to promote to the next grade level. Multiple choice intelligence tests sorted officer candidates from the recruits during WWI. Many of those tests were written and took time to score. Psychologists argued that multiple choice tests were more objective than written responses and therefore more scientifically valid. They were also cheaper and quicker to score, so many educators, beginning in the 1920s, adopted multiple choice standardized tests as the way to evaluate student achievement growth.<sup>24</sup>

The multiple choice tests from this period combined IQ tests with achievement tests, and the results became the mechanism to track students into varying levels of academic programs. Tracking justified stereotypes and created barriers to equal access in school. Another disturbing development occurred with this testing trend; private companies profited greatly from publishing and scoring the tests, so they became agents pushing policymakers to test more.<sup>25</sup> The publishing slowly went from the local level to larger regional levels. The private companies operating from a regional level, and the

---

<sup>24</sup> Diane Ravitch, *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education* (New York: Basic Books, 2010), 151.

<sup>25</sup> Pearl G Solomon, *The Curriculum Bridge: From Standards to Actual Classroom Practice* (Thousand Oaks: Corwin Press, 2009), 151.

administrators' desire for the same operational efficiency they saw in American corporations, caused a further growth in the standardized assessment of students.<sup>26</sup>

Behind this evolution of standardized testing, two pedagogical theories in the early twentieth century defined two very different approaches to education. Edward Lee Thorndike and John Dewey both theorized how humans learn, the purpose of learning, and the role that education should play in society. Thorndike's ideas were instrumental in the standardization movement. He believed that humans are born with a set intelligence. Some people are born to be the leaders in society, and the others are born to be the workers. Thorndike believed in human engineering in that humans are malleable components that can be mechanized and managed. Ordinary people are better off not thinking for themselves. For him, education was very much like running an efficiently operated factory. Society's leaders decide what result they want the students to become. Those decisions shape educational standards. Then, teachers use drill and repetition to make sure the child accomplishes predetermined learning. The multiple choice standardized tests that emerged in the 1920s were the result of the thinking of Thorndike and others who believed that predetermined learning could be identified, defined into a standard, then tested to quantify learning results into measurable numbers.<sup>27</sup> This philosophy about humans connects to the social Darwinist theory, popular at the time, that fueled racism and drove the United States economy and some world events.

---

<sup>26</sup> Sherman Dorn, *Accountability Frankenstein: Understanding and Taming the Monster* (Charlotte: Information Age Publishing, 2007) 33-34.

<sup>27</sup> Stephen Tomlinson, "Edward Lee Thorndike and John Dewey on the Science of Education," *Oxford Review of Education* 23, no. 3 (Sept. 1997), 367-372, <http://www.jstor.org/stable/1050962> (accessed September 6, 2009).

Eventually, that type of thinking was discredited, but unfortunately, it left its mark on education in the standardized testing trend that it helped start.

John Dewey believed that humans can shape their intelligence throughout their lives. Beyond growing their intelligence through experience and inquiry, they have the power and moral right to shape the world around them, including much of the knowledge that defines the way a person understands the world. This idea of shaping knowledge is called constructivism. Beyond constructivism, Dewey believed that all human experience is essentially social and that freedom is ingrained in each person. Educators' responsibility to students was to provide them with opportunities to problem solve. Through cooperative problem solving, students would use their talents to negotiate a solution, much the way citizens should run a democracy. Dewey believed that the mental-testing movement by Thorndike "might become the basis for social stratification and an educational caste system."<sup>28</sup> In contrast to the mental-testing movement, education historians trace the creation of project-based learning back to the ideas of John Dewey.<sup>29</sup>

The standardization movement, started by Thorndike and others who shared his ideas, continued to shape the American public education system for the rest of the twentieth century, but Dewey's ideas, and the ideas of constructivists did not die. Progressive educators and constructivist theorists continued to research and develop ideas similar to Dewey's. Today, many of the modern critics of standardized testing come from a constructivist viewpoint. Their view is that real learning happens with open-ended investigation in which competency is demonstrated through portfolios and

---

<sup>28</sup> Diane Ravitch, *The Revisionists Revised: A Critique of the Radical Attack on the Schools* (New York: Basic Books, Inc., 1978), 136.

<sup>29</sup> Tomlinson, 373-375.

presentations demonstrating project-based learning. Teaching is facilitation, partnership, and collaboration.<sup>30</sup> Constructivists enjoy new common ground with many corporate leaders today who have recently found value in the types of skills that project-based learning teaches.

In the 1930s, dictatorships in Europe and Japan proved to educational leaders in the United States that much more than widespread education of knowledge-based learning was needed to prepare students to become active members of a democracy. Democratic ideals involved practicing skills in school like collaboration, communication, and critical thinking. The Educational Policies Commission (EPC), in a report it published in 1938, reflected these concerns from educational leaders. Consisting of various educational experts and policymakers across the nation like superintendents, scholars, and university presidents, the report from the EPC called for an accountability system that tested more than basic academic skills. The wording from this 1938 report sounds like it came right out of a present day newspaper editorial criticizing the standardized testing movement under *No Child Left Behind*.

Most of the standardized testing instruments and written examinations used in school today deal largely with information....There should be a much greater concern with the development of attitudes, interests, ideals, and habits. To focus tests exclusively on the acquisition and retention of information may recognize objectives of education which are relatively unimportant. Measuring the results of education must be increasingly concerned with such questions as these: Are the children growing in their ability to work together for a common end? Do they show greater skill in collecting and weighing evidence? Are they learning to be fair and tolerant in situations where conflicts arise? Are they sympathetic in the presence of suffering and indignant in the presence of injustice? Do they show greater concern about questions of civic, social, and economic importance? Are they using their spending money wisely? Are they becoming more skillful in doing some useful type of work? Are they living in accordance with the rules of

---

<sup>30</sup> Richard Phelps, "Why Testing Experts Hate Testing," In *Taking Sides: Clashing Views on Controversial Issues in Secondary Education*, ed. Dennis L. Evans (Guilford: McGraw-Hill Companies, Inc., 2002), 157.

health? Are they acquiring the skills in using all of the fundamental tools of learning? Are they curious about the natural world around them? Do they appreciate, each to the fullest degree possible, their rich inheritance in art, literature, and music? Do they balk at being led around by their prejudices?<sup>31</sup>

This EPC report supported John Dewey's ideas of using the classroom to teach students to not only become self-directed learners, but to also become democratic citizens who could work through conflict to reach a consensus that represented the collective values of a democratic society. It also showed a basic assumption from these educational leaders that democracy required people to make it successful, not just leaders that would direct workers to do standardized tasks the way Edward Thorndike envisioned education. This EPC report was the first significant challenge to the standardized testing trend that started in the 1920s.

After WWII, technology and automation defined national priorities. Standardized tests would be used to sort future scientists and industrial managers. A Rockefeller Fund report, *The Pursuit of Excellence: Education and the Future of America*, published in 1958 warned of placing too much emphasis on the results from standardized tests, "Decisions based on test scores must be made with the awareness of the imponderables in human behavior. We cannot measure the rare qualities of character that are a necessary ingredient of great performance. We cannot measure aspiration or purpose. We cannot measure courage, vitality, or determination."<sup>32</sup> This report recognized that there are important skills and traits that are not so easily measured using a standardized test. Too much emphasis on the tests would diminish learning opportunities that teach or foster these important skills or traits. As important as these skills and traits are for students to learn in school, Thomas Jefferson, Horace Mann, and John Dewey all agreed that schools

---

<sup>31</sup> Rothstein, Jacobsen, and Wilder, 24-25.

<sup>32</sup> Rothstein, Jacobsen, and Wilder, 26.

could not fulfill any mission to be “the *sole* instrument of social and economic change.”<sup>33</sup>

The Rockefeller report published in 1958, in the face of a national effort to use standardized testing for technological purposes, demonstrates the ongoing debate and mixed messages over the mission of American education.

A shift occurred in education in the 1960s from localized control to more centralized control. Much of the impetus behind this shift was due to the necessary influence of desegregation policies enforced by the Federal Government over states resisting the 1960s Civil Rights Legislation. However, this shift included more than federal enforcement of equality. It also included using standardized test results as an attempt to monitor the academic growth of students and schools. States started to connect the results with grade promotion or graduation. Until this point, the test scores were used for tracking into vocational or academic programs, which carried with it a whole set of problems, but very few areas in the country used the tests to impede the progress of the student to finish his or her schooling. In the 1960s and 1970s, the test results began to be used for accountability of the students, and then soon after, accountability of the schools. Today, this trend is shifting again, now from accountability of schools to accountability of individual teachers. This latest trend will be explained later in this paper.

From their inception in the early twentieth century to the centralizing shift in the 1960s and 1970s, standardized tests took different forms combining multiple choice, short answer, and essay questions.<sup>34</sup> Even though many of these tests were designed at a regional level, the grading of most of these tests still occurred at a local level where the

---

<sup>33</sup> Diane Ravitch, *The Revisionists Revised: A Critique of the Radical Attack on the Schools*, 10.

<sup>34</sup> Dorn, xvi.

scoring of written responses was more manageable. When these tests became more state level-centrally graded in the 1960s and the 1970s, the short answer and essay questions began to be replaced with all multiple choice questions.<sup>35</sup> Until the 1980s this central control stopped at the state level.

Beginning in the 1980s, the federal government took a more active role in the curriculum of public schools. It started to weigh in on the issue of national standards with its report, *A Nation at Risk*. This report came as recommendations not mandates, and these recommendations were about standards, not about testing. Diane Ravitch, a historian of American education, documented in her book, *Death and Life of the Great American School System: How Testing and Choice Are Undermining Education*, a shift in education that occurred from the early 1980s to the present. She contrasts the report-*A Nation at Risk* with what happened under *No Child Left Behind* (NCLB) legislation in the 2000s:

[*A Nation at Risk*-report] did not refer to market-based competition and choice among schools; it did not suggest restructuring schools or school systems. It said nothing about closing schools, privatization, state takeover of districts, or other heavy-handed forms of accountability. It referred only briefly, almost in passing, to testing. Instead, it addressed problems that were intrinsic to schooling, such as curriculum, graduation requirements, teacher preparation, and the quality of textbooks; it said nothing about the governance or organization of school districts, because these were not seen as causes of low performance.<sup>36</sup>

Federal level-central control over education did not formalize into a system of accountability until the *No Child Left Behind* legislation passed in the summer of 2001. For a short time before this transition occurred, the 1990s saw a renewed interest by various states to use more authentic assessment as portfolios, a testament to the ongoing

---

<sup>35</sup> Solomon, 151.

<sup>36</sup> Diane Ravitch, *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education* (New York: Basic Books, 2010), 25.

debate originally started by Dewey and Thorndike. The 1990s also witnessed a debate over national standards. At this point, states had standards, but they were vague descriptions of what students should know and be able to do. There was an effort by educational reformers to refine the standards to make them more concise and make their effect more meaningful. Unfortunately, the reform efforts went from a standards movement, with authentic assessment as the student achievement measurement, to an accountability movement using standardized testing as the sole mode of assessment. Diane Ravitch described this unfortunate twist as a hijacking of the educational system.<sup>37</sup> The details of *No Child Left Behind* and the evolution of standardized tests that occurred under it will be discussed in further detail in the next section.

In 2004, out of a growing frustration by people who saw that standards-based learning driven by standardized testing was not adequately preparing students with necessary skills, came a goal for learning outcomes called 21st century learning skills. These are skills that involve: communication, problem solving, critical thinking, teamwork, collaboration, cooperation, technology, self-direction, innovation, imagination, creativity, and global awareness.<sup>38</sup> The designers of these learning outcomes involved corporate leaders like Oracle, Cisco, Apple, and Microsoft, and national organizations like the American Association of School Libraries, all of whom recognized certain skills were valuable for the work world, but are difficult to measure on any standardized tests, and are best taught through open learning opportunities like

---

<sup>37</sup> Diane Ravitch, *The Death and Life...*, 15-30.

<sup>38</sup> Some of the strategic council members of Partnership for 21st Century Learning Skills (P21) are: American Association of School Librarians, Education Networks of America, Adobe, Pearson, Apple, Blackboard, Cengage Learning, Cisco, Dell, Houghton Mifflin, HP, Intel, Junior Achievement, McGraw Hill Education, Microsoft, National Education Association, Oracle, Project Management Institute Educational Foundation, Verizon, and the Walt Disney Company. *Partnership for 21st Century Learning Skills*. <http://www.p21.org/> (accessed November 23, 2010).



project-based learning. Today, through a partnership called P21 (Partnership for 21st Century Skills), these corporate leaders and organizations push both basic knowledge and 21st Century Skills.<sup>39</sup> Here is the official statement found on their website:

The Partnership for 21st Century Skills is a national organization that advocates for 21st century readiness for every student. As the United States continues to compete in a global economy that demands innovation, P21 and its members provide tools and resources to help the U.S. education system keep up by fusing the three Rs and four Cs (critical thinking and problem solving, communication, collaboration, and creativity and innovation). While leading districts and schools are already doing this, P21 advocates for local, state and federal policies that support this approach for every school.<sup>40</sup>

The skills that P21 encourages educators to teach cannot easily be measured on standardized tests. Potentially, this contradicts with the TAP system that uses standardized test scores to evaluate teachers, depending on how the local district decides which assessments will be used to evaluate student achievement growth. As previously mentioned, the majority of TAP schools choose to use scores from standardized tests to evaluate their teachers, because standardized tests are already an established assessment mechanism with high-stakes ramifications attached, and the results are easy to compute.

The Federal Government and Sycamore River endorse both TAP and P21, which, when analyzed closely is a fundamental disconnect between education policy and educational research. How did this disconnect develop? Skills needed for the 21st century are best taught and practiced with project-based learning, but the reality is that standardized tests are pushed onto local schools by the state and national governments. This all occurred while educational research found that these tests were not accurate or

---

<sup>39</sup> Bob Pearlman, "21st Century Learning in Schools--A Case Study of New Technology High School in Napa, CA." <http://www.bobpearlman.org/Articles/21stCenturyLearning.htm> (accessed May 12, 2011), 1.

<sup>40</sup> Partnership for 21st Century Learning Skills. <http://www.p21.org/> (accessed November 23, 2010).

authentic. Society blamed teachers for the perceived ills within education, because students were not leaving school prepared, and corporations were frustrated that workers were not prepared to handle the work-world, so they pushed for a set of skills. However, this set of skills cannot be measured on standardized tests. For some reason, though, these corporate leaders would not speak out against standardized testing. Do the supporters that push the agenda to privatize public schools want to use high-stakes, standardized, accountability to weaken society's perception of public schools? The answers to these questions are beyond the scope of this paper, but they are worth considering after this paper unveils the magnitude of the historical disconnect that is being described here.

Today, instead of shaping policy around the research findings, policymakers push for more standardized testing and testing results are connected to teacher performance pay. The idea goes something like this: *if teachers could just be motivated to work harder, then our students' scores on their standardized tests would go up, and they would be more ready for the competitive global world we live in today.* The result of this disconnect between educational research and education policy is that teachers, instead of investing time and energy to facilitate purposeful project-based learning, teach tested facts.

Take for instance the requirement that my fifth grade son memorize the capital of each state in the United States. Why? This information that can easily be looked up on his *iTouch*, and is arguably not important information. Somewhere along the line someone decided that it would be important for little boys and girls in the U.S. to learn the capitals of each state, and no one challenges whether the exercise is an inefficient use

of learning time. Why is Albany, New York worth remembering more than New York, New York? Who cares if Cheyenne is the capital of Wyoming, unless of course you are one of the 563,000 people that live in Wyoming. The city of Fresno has more people. Sure, each of those capital cities have historical significance, but the students are not learning the history of each state. If the argument is that they should practice memorization skills, then surely, there are more valuable things to memorize like math facts that are foundational to math computation.

Instead of memorizing the state capitals, they should learn the twenty most populated cities in the nation and how they became so populated. Then, more meaningful learning could incorporate historical lessons in immigration, commerce, transportation, environmental concerns, infrastructure costs, crime and other effects of population, with present day business and politics. The lessons could be split into project-based learning with each group presenting their findings, connecting history with present day lessons.

In Robert Frost's poem "Mending Wall," the narrator challenges *tradition* by asking himself why do "good fences make good neighbors" when he and his neighbor don't need the wall to fence out cows. He only has apple trees and his neighbor only has pine trees, but they mend the wall each spring because it is an idea passed down from the previous generation. Interestingly, the narrator thinks about how the tradition makes no sense for him and his neighbor, and yet he continues the *tradition*.

The *tradition* we face deals with the way we spend our classroom minutes. Our history delivered us to this present moment with valuable lessons learned. Educational research findings in the past revealed that standardized test results are inadequate tools for measuring student achievement. By looking at the past it becomes clear that teacher

merit pay systems have a strong record of failure. Although history is not normative, a reasoned approach toward the future warns that teacher performance-pay based on standardized testing is doomed from the start, because it wraps two bad ideas into one.

### Chapter III The Inadequacy of Standardized Tests

It is easy to see why policymakers mandate standardized testing; they are easy to administer and to score. The results translate into tidy percentages that impress parents and the media with the perceived comparability they offer between students, teachers, schools, districts, and states, and even nations. They seem like a concrete way to evaluate teachers, schools, and districts. Society hungers for accountability and competition.<sup>41</sup> Corporations profit from selling tests and other services that accompany testing. Test results will never live up to *No Child Left Behind* standards, so the results offer ammunition for the privatization movement that wants to dismantle public education and the teachers' unions. The engraining of standardized tests in our culture reaches back almost a century, so it would be considered Un-American if we got rid of them. They also offer a mechanism of control from federal and state governments over local districts.

Educational policy continues to support standardized testing for many of the reasons listed, but those reasons simply do not hold up to the scrutiny of educational research that has studied the issues and complexities around standardized testing. For decades, it has been well known from educational research findings that standardized tests are not accurate or authentic. The "Survey of Initial Sources," section explains how a majority of sources argue against standardized testing. The remainder of this section provides samples and commentary regarding the sources dealing with standardized testing.

---

<sup>41</sup> Herndon, Terry. "Merit Pay and the Concerns of the Teaching Profession." In *Merit, Money and Teachers' Careers: Studies on Merit Pay and Career Ladders for Teachers*, edited by Henry C. Johnson, Jr., 93-98. (Lanham: University Press of America, Inc., 1985), 93.

Peter Sacks attacked the validity of standardized tests in a 1997 article when many states stood at a crossroad regarding standardized testing and alternative accountability measures. In 1980, only 26 states mandated testing programs, but by the time Sacks published his study, 46 states mandated testing. In the early nineties, as the trend to adopt mandated testing among the states grew, some education researchers chose to revisit the idea of accountability and consider if there were alternatives. The government responded to that desire. In 1992, the United States Office of Technology Assessment reported, “It now appears that the use of these tests misled policymakers and the public about the progress of students, and in many places hindered the implementation of genuine school reforms.”<sup>42</sup> Sacks’ article listed several reasons why testing blocked true reform: tests are costly; they do not accurately measure student achievement; scores correlate to socioeconomic class; and standardized tests drive instruction in the wrong direction, toward knowledge-based, superficial learning rather than higher level thinking skills. According to Sacks, America’s fascination with test scores is unique compared to other countries, like in Europe, where more authentic measurements like portfolios or essays demonstrate higher level knowledge. Finland boasts some of the highest achievement of students in the world. One article explained two reasons why Finland reached this level of achievement:

The first is Finnish citizens held teachers and school principals in the highest esteem. Those who graduate at the top of their class are the only ones who can consider a career in education. It is the most competitive field, more so than medicine and law. The average acceptance rate into schools of education is a mere 10%.

---

<sup>42</sup> Peter Sacks, “Standardized Testing: Meritocracy’s Crooked Yardstick,” in *Taking Sides: Clashing Views on Controversial Issues in Secondary Education*, ed. Dennis L. Evans (Guilford: McGraw-Hill Companies, Inc., 2002), 172.

The second thing happened in the 1980s: Finland abolished standardized tests. Instead of test-based accountability in schools, the country—because of the high quality of its teaching force—had a trust-based system to allow teachers a certain freedom to teach with creativity. Students, too, had autonomy to learn in different ways....Sahlberg [the director general of the Center for International Mobility and Cooperation in Finland] argued that fewer teaching hours means more time for educators to create interesting lessons, to apply authentic assessments, and grant students liberal time to work on their studies and projects.<sup>43</sup>

Finland abolished its own standardized tests, and yet scores well on international tests that are also standardized. It is unclear why Finland scores so well on international standardized tests like the TIMSS (Trends in International Mathematics and Science Study). Perhaps, it is because they have embraced project-based learning. Perhaps, it is because nearly all Finnish students speak the same language as their teachers. Whereas in the United States, with its high level of immigration, one out of five students speak another language other than English at home. Jared Graham, in his article, “Finland test scores in perspective,” suggests that test scores between nations with different populations is not comparable.<sup>44</sup>

According to Sacks, there needs to be an American cultural shift away from its obsession with standardized tests and the results from such tests. In the early 1990s, twenty-one states had launched performance assessment programs. According to Sacks, a conservative backlash caused some of these states to reconsider their retreat from standardized testing. California, Arizona, and Indiana all decided to stop any move

---

<sup>43</sup> *Asia Society* (author unknown), “What Accounts for Finland’s High Student Achievement Rate?” *Asia Society*, April 27, 2010. <http://asiasociety.org/education-learning/learning-world/what-accounts-finlands-high-student-achievement-rate> (accessed June 25, 2011).

For another discussion of education in Finland, see: Caroline V. Gipps, “Accountability Testing and the Implications for Teacher Professionalism,” in *Measurement and Research in the Accountability Era*, ed. Carol Anne Dwyer (Mahwah: Lawrence Erlbaum Associates, Publishers, 2005) 99-105.

<sup>44</sup> Jared Graham, “Finland test scores in perspective,” *Education in Practice: Exploring Issues in Education and Education Reporting*, August 12, 2010, <http://www.educationinpractice.net/tag/international-test/> (accessed July 8, 2011).

toward alternative forms of assessment and revisit implementation of high-stakes standardized testing.

Many of these alternative assessments involved writing essays and short answers. This form of assessment requires real people to evaluate, and therefore costs more money than many states are willing to pay, if they are evaluated on a statewide level. However, if these assessments are evaluated on a local level, it wouldn't cost the public anything, because teachers on a local level could evaluate them as part of their job. The accuracy of the local evaluations could be monitored by state level representatives, much like a visiting accreditation committee evaluates a school, or a state level regulatory group could double score a small percentage of local assessments to check local scoring for inter-rater reliability. Perhaps even the state level monitoring of local evaluations would be considered too costly, but the only cost that should be considered is not the cost of evaluations, but the cost of meaningful learning. When considering the fact that alternative ways to assess exist, why do policymakers stick with standardized testing?

Sacks addressed one of those reasons. Some defense of standardized testing came with the challenge to affirmative action at the university level. To avoid overusing ethnicity and gender in the decision making, college admission boards decided to use "merit" as a way to rank college applicants. The easiest way to establish that ranking was the use of standardized test scores. An increased need to rank college applicants came with the increased population levels from the baby boom. This population increase directly correlates with increased use of standardized testing, which saw a sharp spike in the 1960s. Between 1960 and 1989, the sale of standardized tests (after adjusting for inflation) more than doubled. Sacks credited a conservative movement during this time



to use standardized testing to measure “merit.”<sup>45</sup> The problem, according to Sacks, is that standardized test scores are highly correlated with socioeconomic class. He called it the “Volvo Effect.” A 1991 study backed Sacks assertion by showing that SAT results show a trend of testers scoring thirty points higher for every \$10,000 in parents’ yearly income. In 1995, the U.S. Department of Education reported only eight percent of students who made the SAT minimum cut of 1100 points for highly selective colleges came from lower economic rungs, and one-third came from upper income brackets.<sup>46</sup> With the choice to use standardized testing as a tool to sort students into categories, policymakers sacrificed an emphasis on important skills in favor of a standardized format easily scored by a computer. Sacks quoted an educational researcher, Bruce C. Bowers, who stated that short answer, multiple choice tests lowers the priority status of skills such as writing, speaking, acting, drawing, constructing, and repairing.<sup>47</sup> Why can’t universities trust local level teacher evaluators with statewide monitoring to assess skills such as these and consider college applicants for the qualities that universities truly desire?

One of Sacks three points was “Standardized tests generally have questionable ability to predict one’s academic success, especially for certain subgroups.” Sacks used a 1994 study by Educational Policy to show the viewpoint of teachers regarding mandated standardized testing. Only three percent of teachers in one sample agreed the tests are good, “whereas 77 percent felt that tests are bad and not worth the time and money spent on them,” and eight out of ten of the teachers believed that their colleagues taught to the

---

<sup>45</sup> Sacks, 168-169.

<sup>46</sup> Sacks, 171.

<sup>47</sup> Sacks, 172.

test.<sup>48</sup> A study in 1995 regarding the Graduate Record Exam (GRE) supported the majority opinion of teachers who felt the tests were bad. In that study, the educational performance of 5,000 GRE test-takers (over the course of thirty years) showed that “GRE scores accounted for just 6 percent of variation in grades in graduate school.” The authors of the study concluded that the GRE appears to be “virtually useless from a prediction standpoint.”<sup>49</sup> Sacks concluded this part of his argument with the idea that standardized test scores are only good for predicting how well someone performs on standardized tests. Those students entering medical school, law school, or teaching can look at their standardized test scores as a predictor of performance on future standardized tests within their field of study, but for everyone else, standardized test scores do not predict academic performance beyond the first year of college or in the workplace.<sup>50</sup> It is worth questioning the validity of standardized tests even within the medical, law, and teaching fields. Why don’t these fields also place more emphasis on authentic assessment? How ironic that the teaching field uses standardized testing to field its future candidates when the survey results of teachers regarding standardized testing shows such disfavor for that type of assessment. However, recently a trend to use authentic assessment to evaluate teacher candidates grows. California leads the nation with this type of authentic assessment of its teacher candidates, known as PACT (Performance Assessment for California Teachers), which avoids a *one size fits all* approach and allows each candidate to demonstrate competency in an open-ended, flexible manner.<sup>51</sup>

---

<sup>48</sup> Sacks, 172.

<sup>49</sup> Sacks, 170.

<sup>50</sup> Sacks, 171.

<sup>51</sup> PACT was designed to prevent “bad teachers” from ever obtaining a teaching credential. PACT (Performance Assessment for California Teachers), <http://www.pacttpa.org/main/hub.php?pageName=Home> (accessed June 23, 2011). This is a description from the PACT website, “Among other provisions, it established a requirement for all

Arguments made by supporters of standardized testing continue to fuel the debate against an overwhelming amount of research that finds serious faults in standardized testing. By comparing the ratio of research findings in my “Survey of Initial Sources,” it is evident that proponents of standardized testing are in the slim minority. Richard Phelps is one of those supporters of standardized testing. His article is featured in *Taking Sides: Clashing Views on Controversial Issues in Secondary Education*, opposite to Sacks argument. Phelps accurately recognized that constructivists are the main critics of standardized tests, because the tests ask questions with prescribed answers, and limit the ability of the student to make knowledge.

To argue against constructivism, Phelps used the example of a surgeon requiring a considerable amount of rote memorization, routine, and factual recall to do his job, all of which constructivists would call lower order thinking. Unfortunately, the debate between supporters of standardized tests and constructivists can be so polarized that one is left with an either/or choice. Phelps led the reader to assume that constructivists do not allow for direct instruction, which is true of some extremist supporters of constructivism, but not true of the constructivist who supports project-based learning. For them, it is not a choice between only rote knowledge and the choice between only constructing knowledge. Surgeons, for example, require an extensive foundational knowledge base and the ability to think critically. Residency, following medical school, is a perfect example of scaffolding the learning from base knowledge to real world application. Phelps portrayed a constructivist as someone who detests any form of base knowledge.

---

California candidates for a preliminary teaching credential to pass a state-approved teaching performance assessment with demonstrated validity and reliability to supplement training, course assignments and supervisor evaluations. The California Commission on Teacher Credentialing contracted with the Educational Testing Service to develop such an assessment.”

Considering my research, the constructivist thinking behind project-based learning recognizes the need for base knowledge, but sees more value in learning that base knowledge while applying critical thinking to solve a problem. Within project-based learning, direct instruction occurs, but it is strategically timed with teachable moments when the students need direct instruction to make progress on their project. Immediately the information has an applied purpose. The results show that project-based learning does a better job building base knowledge, according to the experiences of New Tech High, Central Park East, and the Buck Institute for Education.<sup>52</sup>

According to Phelps, “There is no necessary correlation between the difficulty of a problem and its response format,” so a problem that would require fifty minutes to “classify, assemble, organize, calculate, and analyze” can still present a multiple-choice response format.<sup>53</sup> He was correct in recognizing that higher level problem solving can result in a multiple choice response. However, the multiple choice questions that appear on the standardized testing in public schools do not require fifty minutes to “classify, assemble, organize, calculate, and analyze.” That much time required for one question discourages accuracy, especially from students who figure that the results from these tests have no direct bearing on their lives, and in actuality only affect teachers and administrators. In reality, the questions posed to students on standardized tests are allotted around sixty-seconds of response time. In a testing period, usually four to eight

---

<sup>52</sup> Manor New Tech High School. <http://www.manorisd.net/portal/newtech/> (accessed June 17, 2011).

New Tech Network. <http://www.newtechnetwork.org/> (accessed June 17, 2011).

Buck Institute for Education: <http://www.bie.org/> (accessed January 20, 2011).

Deborah Meier, *The Power of Their Ideas: Lessons for America from a Small School in Harlem* (Boston: Beacon Press, 2002).

<sup>53</sup> Richard Phelps, “Why Testing Experts Hate Testing,” In *Taking Sides: Clashing Views on Controversial Issues in Secondary Education*, ed. Dennis L. Evans (Guilford: McGraw-Hill Companies, Inc., 2002), 161.

part-time days, students will answer approximately 300-500 questions. With an exception to math and some English language arts, most of these questions require rote memorization to successfully answer.

According to Phelps, teachers in America that oppose standardized testing should take a lesson from teachers in other countries who embrace standards. Phelps claimed that this “standardization brings the security, convenience, camaraderie, and common professional development that accompany a shared work experience.”<sup>54</sup> He mistakenly classified standards and standardized testing as the same thing. Phelps criticized those teachers that oppose standardized testing by assuming that they are opposed to standards. Standards are merely a common set of goals of what students should know and be able to do and attainment of complex goals can be measured accurately with authentic assessment.<sup>55</sup>

---

<sup>54</sup> Phelps, 160.

<sup>55</sup> Betty E. Steffy, and Fenwick W. English, *Curriculum and Assessment for World-Class Schools* (Lancaster: Technomic Publishing Company, Inc., 1997).

Linda Darling-Hammond, Jacqueline Aness, and Beverly Falk, *Authentic Assessment in Action: Studies of Schools and Students at Work* (New York: Teachers College Press, 1995).

Daniel Schwartz, Robb Lindgren, and Sarah Lewis. “Constructivism in an Age of Non-Constructivist Assessments,” In *Constructivist Instruction: Success or Failure*, ed. Sigmund Tobias and Thomas M. Duffy (New York: Routledge, 2009), 34-61.

Carol Anne Dwyer, ed., *The Future of Assessment: Shaping Teaching and Learning* (New York: Lawrence Erlbaum Associates, Publishers, 2008).

Mark Wilson, ed., *Towards Coherence Between Classroom Assessment and Accountability: 103rd Yearbook of the National Society for the Study of Education, Part II* (Chicago: The University of Chicago Press, 2004).

Robert L. Linn and M. David Miller, *Measurement and Assessment in Teaching*. 9th ed. (Upper Saddle River: Pearson Education, Inc., 2005).

Alfie Kohn, *The Case Against Standardized Testing: Raising the Scores, Ruining the Schools* (Portsmouth: Heinemann, 2000).

Mary Ellen Freeley and Richard Hanzelka, “Getting Away from Seat Time: A New Hampshire initiative encourages schools to move toward competency-based learning,” *Educational Leadership* 67, no. 3 (November, 2009): 63-67.

Pat Belanoff and Marcia Dickson, eds., *Portfolios: Process and Product* (Portsmouth: Boynton/Cook Publishers, Inc., 1991).

Deborah Meier, *The Power of Their Ideas: Lessons for America from a Small School in Harlem* (Boston: Beacon Press, 2002).

Manor New Tech High School. <http://www.manorisd.net/portal/newtech/> (accessed June 17, 2011).

The problem with standardized testing is the way only some of those standards are allowed to be tested, because of the limiting factor of multiple choice responses. Diane Ravitch very carefully separated the support for standardized testing and standards. She wrote, “Accountability makes no sense when it undermines the larger goals of education....Somehow our nation got off track in its efforts to improve education. What once was the standards movement was replaced by the accountability movement.”<sup>56</sup> There is nothing wrong with society agreeing upon a base set of standards, but the critical question that should drive policy is the ultimate goal for our students to achieve. Multiple choice scores that serve as the endgame distracts from the real-life needs of students to be able to write, speak, act, draw, construct, repair, evaluate, predict, analyze, and connect. The distinction between standards and standardized testing cannot be overstated. Diane Ravitch learned this lesson the hard way.

Ravitch brings a unique voice to this debate over standardized testing. As an historian of American education, she understands the historical affect from project-based learning, standards, and standardized testing in the United States. She is one of the original leading proponents of *No Child Left Behind* (NCLB). At its inception, she mistakenly thought that NCLB would launch a national standards movement to encourage equal access to knowledge and skills for children across the United States. As the former Assistant Secretary of Education under President George W. Bush, she pushed for the creation of NCLB, and recently she explained in her book why she is no longer supporting NCLB. For her, it started out as a standards movement and turned into an

---

New Tech Network. <http://www.newtechnetwork.org/> (accessed June 17, 2011).  
 Buck Institute for Education: <http://www.bie.org/> (accessed January 20, 2011).

<sup>56</sup> Diane Ravitch, *The Death and Life...*, 16.

accountability movement that is not connected to the knowledge and skills necessary for success.<sup>57</sup>

For Ravitch, the debate over standardized testing is about the relationship between democracy and schools and what it takes to create a civil society. This idea reaches back to John Dewey and his ideas about how project-based learning is the most effective way to teach students how to be valuable members of a democracy, and how it equips students to be able to reach their potential. Ravitch discussed how standardized testing is taking education in the U.S. in the wrong direction by turning education into a marketplace of consumers rather than a series of smaller neighborhood schools built around a community that looks out for its own students. Standardized testing can push educators to leave children behind, quite opposite to its original intent.

The Texas standardized testing movement, that served as the model for how Congress designed NCLB, showed test scores going up at the cost of African American and Hispanic dropout rates also going up. In separate studies, Boston College and RAND both concluded that Texas State test gains did not lead to gains in SAT, National Assessment of Educational Progress (NAEP), or even Texas college readiness scores.<sup>58</sup> The study from Boston College found that curriculum in Texas was narrowed to drive up test scores, and that “such subjects as science, social studies, and the arts were pushed aside to make time for test preparation. Consequently, students in Texas were actually

---

<sup>57</sup> Diane Ravitch, *The Death and Life...*, 28-30. It is important to note at this point in the paper, that the focus of this paper is not exclusively on NCLB, because the standardized testing movement in this country goes beyond NCLB. However, beginning with its enactment in the summer of 2001, NCLB became the dominant force behind standardized testing. Now that the tests from NCLB are in place and carry such weight within people’s perception of schools, teacher incentive programs have been created that are directly connected with the standardized tests from NCLB.

<sup>58</sup> For an explanation of NAEP, see The National Center for Education Statistics <http://nces.ed.gov/nationsreportcard/about/> (accessed June 24, 2011).

getting a worse education tied solely to taking the state tests.”<sup>59</sup> Unfortunately, the creators of NCLB did not look hard enough at these results of the Texas standardized testing movement. At the inception of NCLB Ravitch admitted that:

I too had fallen for the latest panaceas and miracle cures; I too had drunk deeply of the elixir that promised a quick fix to intractable problems. I too had jumped aboard a bandwagon, one festooned with banners celebrating the power of accountability, incentives, and markets. I too was captivated by these ideas. They promised to end bureaucracy, to ensure that poor children were not neglected, to empower poor parents, to enable poor children to escape failing schools, and to close the achievement gap between rich and poor, black and white. Testing would shine a spotlight on low-performing schools, and choice would create opportunities for poor kids to leave for better schools. All of this seemed to make sense, but there was little empirical evidence, just promise and hope.<sup>60</sup>

Ravitch supported NCLB until November 30, 2006, when she attended a conference and heard a dozen research experts present their analyses of NCLB’s remedies.<sup>61</sup> The conference focused mostly on the effectiveness of two of NCLB’s strategies for educational improvement--choice and after school tutoring. They all agreed that choice was not working. Under NCLB rules, if schools failed to meet their designated Annual Yearly Progress (AYP) a second year in a row, then students would have the choice to transfer to a successful school, with transportation paid for by the district through allotted federal funds. The scholars’ findings showed that parents and students were not taking advantage of the choice to move schools.<sup>62</sup> There were many

---

<sup>59</sup> Diane Ravitch, *The Death and Life...*, 96.

<sup>60</sup> Diane Ravitch, *The Death and Life...*, 3-4.

<sup>61</sup> It is worth pointing out that this conference was put on by the American Enterprise Institute in Washington, D.C.--a well respected conservative think tank. If this group was driven by any sort of bias at all, then it would have been toward supporting NCLB which President Bush toted as the way to reform education. In fact, the scholars put together by the American Enterprise Institute showed that NCLB was not working.

<sup>62</sup> Here are the states listed by Ravitch as examples: California, less than 1 percent of eligible students in “failing” schools sought to transfer to another school; in Colorado, less than 2 percent; in Michigan, the number was negligible, in Miami, where public school choice was already commonplace, less than 1/2 of 1 percent asked to move because of NCLB; in New Jersey, almost no eligible students transferred, page 99.



reasons why students did not move when given a choice: desire to go to a neighborhood school, not wanting to ride a bus across town, not wanting to leave a successful school when it had been labeled “failing,” because only one subgroup--usually children with disabilities--did not make adequate progress. The scholars’ analyses also revealed that only twenty percent of students eligible for free tutoring took advantage it. Billions of dollars went to companies that offered tutoring, tests, and test prep materials. After looking closely at these issues, Ravitch admitted, “The advantages to the nation’s students were not obvious.”<sup>63</sup>

She concluded her turnaround lesson on NCLB by summarizing the realization she had that day at the conference. According to Ravitch, the federal government should supply valid information and give more power to decide remedies and sanctions to those closest to the problems. The incentives and sanctions set up by NCLB are appropriate for business goals of profit, but not education. The “Proficiency” goal (which shows high academic ability) of 100% of students in America by 2014, which includes students with special needs, students whose first language is not English, and students who are homeless, is not realistic and will have a privatizing effect, according to Ravitch. She admitted that members of Congress probably did not realize the way public education could fundamentally change if the NCLB mandates are enforced in 2014. She wrote, “I started to see the danger of the culture of testing that was spreading through every school in every community, town, city, and state. I began to question ideas that I once embraced, such as choice and accountability, that were central to NCLB. As time went by, my doubts multiplied.”<sup>64</sup>

---

<sup>63</sup> Diane Ravitch, *The Death and Life...*, 100-101.

<sup>64</sup> Diane Ravitch, *The Death and Life...*, 102.

How can the public come to the same realization that Ravitch came to that day, November 30, 2006, when the members of the public are not as knowledgeable about the issues and details as Ravitch or the research experts who spent much time and energy analyzing the details? Diane Ravitch, whose job at the time was Assistant Secretary of Education, who is an author of multiple books dealing with the history of education in America, did not realize that NCLB was failing until she attended a conference in Washington D.C. with experts presenting their analyses, six years after the inception of NCLB, and after the red flags had already well documented the Texas standardized testing movement. The public cannot realize the failings of NCLB or the standardized testing movement on a national level unless the experts show them the same results that Ravitch saw that day at the conference. Many parents are experts over their individual situations. They don't need a standardized test to tell them whether a teacher or a school is making adequate yearly progress with their child. If the public believed in and supported the mandates set up by NCLB, then why didn't more parents take advantage of transferring from "failing" schools, or why didn't more parents take advantage of free tutoring offered by the schools?

Richard Phelps published his article in 1999 at the same time that Diane Ravitch was, in her own words, "jump[ing] aboard a bandwagon, one festooned with banners celebrating the power of accountability, incentives, and markets." Phelps argued in his article that there was overwhelming public support for high-stakes standardized testing. After searching Roper Center archives, he discovered "200 items from seventy-five surveys over three decades" that show majorities of the general public, including parents, student, employers, state education administrators, and even teachers who "consistently

favor more student testing and higher stakes.”<sup>65</sup> If Diane Ravitch, an “insider” nearest to policymaking, was misled by the promises of high-stakes standardized testing, then wouldn’t most of the public also be misled? Phelps argued the merits of standardized testing by showing that the majority of the public supported it, but as we know from history, the majority opinion can be wrong.<sup>66</sup> There can exist a disconnect between what the public thinks and what research findings show. Instead of using public surveys to prove that standardized testing is authentic and accurate, it is more valuable to look at what educational research has concluded over the years.

What did educational research say about standardized testing several years before NCLB was passed by Congress? The Educational Research Services, known as *The Information Source for School Decisions*, is sponsored by the American Association of School Administrators and six other educational policy organizations. This group of experts published an article in 1997, “What Do We Know About Declining (or Rising) Student Achievement?” In this article, Richard Rothstein argued that student achievement in the past cannot be compared to recent student achievement for six reasons: tests change, curricula change, student populations change, unreliable background data, and inconsistent test administration. The few measures that have been around the longest show a slight increase in student performance. He explained that three assessments (SAT, ITBS, NAEP) all show no significant decline of student achievement. The NAEP pattern of *increase* of student achievement among white students continued to

---

<sup>65</sup> Phelps, 165.

<sup>66</sup> Many Americans supported going to war in Iraq after 911, because policymakers told them it was a good idea due to the idea that Iraq was directly connected to 911, and that Iraq was becoming a nuclear threat. In this case, the information from intelligence research, which stood in direct contrast to the policymakers’ message, did not get to the public. The policymakers used misinformation to achieve their agenda. In the case of high-stakes standardized testing, it is best to analyze the merits of the tests by looking at what educational research concludes, not what the public is saying in response to policymakers message driven by agenda.

occur and surprisingly, minority students' *scores improved* more than predicted.<sup>67</sup>

Rothstein's main point was that the perception that the educational system in the U.S. was broken--was a false perception.

A year before the president signed NCLB into law, Alfie Kohn wrote an article and a book arguing against standardized testing. He is a former educator who now writes about education and was recognized by *Time* magazine, at the same time that he published his article and book, as "perhaps the country's most outspoken critic of education's fixation on grades [and] test scores." His book compiled findings from 123 experts. It is a quick read with only 66 pages, followed by 12 pages of notes, and 22 sources for suggested further reading. Members of Congress who voted for NCLB, and Diane Ravitch, would have known about his publications and the experts he references, if they had looked for expert advice. In total, Kohn has authored eight books on education and human behavior. His perspective is that standardized tests scores tell us very little about student learning, and that curriculum driven by standardized tests are tragically missing the opportunity to teach students to think their way through a real life problem. His basic question asked, "How can parents be confident that their child is learning."<sup>68</sup> His answer stated a need to create clearly identified competencies and portfolios with teacher narratives to accompany the portfolio to show whether the student has met that competency.<sup>69</sup>

---

<sup>67</sup> Richard Rothstein, *What Do We Know About Declining (or Rising) Student Achievement* (Arlington: Educational Research Service, 1997), i-ii.

<sup>68</sup> Alfie Kohn, *The Case Against Standardized Testing: Raising the Scores, Ruining the Schools* (Portsmouth: Heinemann, 2000), 41.

<sup>69</sup> This idea is apart of a larger debate that occurred in the 1990s between using portfolios or using standardized testing as the method for accountability. That debate seemed to end with the passage of NCLB. Despite the fact that the experts continued to argue against standardized testing, the policy-makers behaved as if the problem of accountability had been solved. The more time that passes from that debate in the 1990s about portfolio assessment seems to

Kohn is a constructivist who credited some of his ideas to Dewey and Piaget. For him, learning is about actively making meaning, which results in higher level thinking. The tests reflect low level thinking. He used findings from Lauren Resnick, a leading cognitive scientist, to illustrate his argument against standardized testing. Most standardized testing questions test information that is crammed into short-term memory. Even the rare questions that require reasoning generally fail “to carry out extended analysis, to solve open-ended problems, or to display command of complex relationships, although these abilities are at the heart of higher-order competence.”<sup>70</sup> The questions that supposedly test reading comprehension “rarely examine how students interrelate parts of the text and do not require justifications that support the interpretation,” which leads to “the quick finding of answers rather than reflective interpretations.”<sup>71</sup> Kohn stated that questions dealing with social studies and science do not test whether the student knows to think like a scientist or a historian, but instead the tests are designed “to tell who can recite the four stages of mitosis or the four freedoms mentioned by Franklin Roosevelt.”<sup>72</sup> Proficiency on the test often comes down to memorization rather than critical thinking. One study that focused on fifth graders and their understanding of division as reflected in standardized test responses showed that true understanding of division among half the students would have been misclassified, because “41 percent had memorized the process without really understanding the idea, while 11 percent

---

correlate with the size of the disconnect between policymakers and experts over the issue of accuracy and authenticity of standardized tests.

<sup>70</sup> Kohn, 7. This quote from Lauren Resnick is from her book, *Education and Learning to Think*. Washington, D.C., National Academy Press, 1987, 34.

<sup>71</sup> Kohn, 8. The quotation here is actually Kohn quoting from Lauren Resnick and Daniel Resnick in *The Uses of Standardized Tests in American Education: Proceedings of the 1989 ETS Invitational Conference*. Princeton, N.J.: Education Testing Service, 1990.

<sup>72</sup> Kohn, 9.

understood the concept but made minor errors that resulted in getting the wrong answers.”<sup>73</sup>

In the fall of 2000, as NCLB was being pushed through Congress, Kohn wrote an article as a desperate plea for people to wake up to the danger of the standardized testing movement. He wrote, “As the year 2001 begins, we are facing an educational emergency in this country. The intellectual life is being squeezed out of schools as they are transformed into what are essentially giant test-prep centers.”<sup>74</sup> He provided a list of issues related to learning that he felt would be seriously harmed as long as standardized testing is allowed to dominate the educational agendas of policymakers and administrators. These issues are: multiple intelligences, multiage classrooms, or multicultural curricula; cooperative learning, character education, or the creation of caring communities in schools; teaching for understanding, developmentally appropriate practice, or alternative assessment; the integrations of writing or the arts into the curriculum; project-or problem-based learning, discovery-oriented science, or whole language; giving teachers or students more autonomy, or working with administrators to help them make lasting change.<sup>75</sup> One goal of this paper is to show what has occurred in Sycamore River Unified School District as a result of the emphasis placed on standardized testing since 2001. Many of Kohn’s warnings have come to fruition in Sycamore River. The last section of this paper fully discusses these issues.

---

<sup>73</sup> Kohn, 11. The findings from this study are reported in “Improving Instruction Through Brief Interviews.” *Arithmetic Teacher*, November 1989: 15-17.

<sup>74</sup> Alfie Kohn, “Fighting the Tests: A Practical Guide to Rescuing Our Schools,” *Cultural Logic*, Vol. 4, Number 1, (Fall, 2000) <http://clogic.eserver.org/4-1/kohn.html> (accessed 10/31/2010), 2.

<sup>75</sup> Alfie Kohn, “Fighting the Tests: A Practical Guide to Rescuing Our Schools,” 2.

In July 2001, a month after the Senate passed NCLB, the Educational Testing Services (ETS), the world's largest private nonprofit educational testing and assessment organization, published a study that brought together the findings of eight publications of the ETS Policy Information Center as well as numerous other reports and sources. The ETS published the study to warn policy-makers that testing is not the end-all for educational reform. Paul Barton, the author of the report, wrote, “Standards-based reform is in danger of becoming simply a testing movement; testing itself is not the treatment, but a way of finding out whether new content standards, rigorous curriculum, and teacher preparation are producing results. In a full standards-based reform effort, testing is just one important component.”<sup>76</sup> This is exactly why Diane Ravitch, five years later, withdrew her support for NCLB. In the booklet, Barton synthesized a decade of findings from the ETS that showed how the validity of tests can be distorted depending on the behavior of the students. “Some of these behaviors have obvious impacts on teaching and learning. Class cutting, absenteeism, and tardiness reduce learning time. Drugs and alcohol are a drag on health, on studying, on attention, and on attendance.”<sup>77</sup> These behaviors are outside the control of individual teachers, and yet the solution that society demands is that teachers just be more accountable. Barton warned educators to use multiple criteria for high-stakes decisions. If teacher performance pay is used, then it should be implemented with multiple measures, with standardized test scores serving as a one small measure.

Despite Barton’s warnings, by 2002, nearly every state had implemented some type of incentive connected to the scores from high-stakes standardized testing. The amount

---

<sup>76</sup> Paul E. Barton, *A Policy Information Perspective: Facing the Hard Facts in Education Reform* (Princeton: Educational Testing Service, 2001), 30.

<sup>77</sup> Barton, *A Policy Information Perspective: Facing the Hard Facts in Education Reform*, 12.

of money drew the attention of economic researchers who noticed that in California alone, in 2001, \$700 million was spent on this type of financial incentive. In an article titled, “The Promise and Pitfalls of Using Imprecise School Accountability Measures,” Thomas Kane and Douglas Staiger, two economists, showed how test score measures can be imprecise and therefore less reliable than most people realize. The authors found that some states allow certain types of student testing exemption, like students with learning disabilities, limited English proficiency or who are absent on the day of the test, which gives “school personnel considerable opportunity to manipulate which students take the test and thus affect average performance.”<sup>78</sup> Ways to manipulate the test results, non-optimal test days (dog barking outside, gloomy weather, tester having a bad day), single year averages, are all examples of how the test results may not reflect the actual learning that is occurring. Financial awards varied state to state, with Texas receiving less than \$5,000 per school and California receiving as high as \$50,000 per school. The article made an important economic policy point that money used to fund incentive programs could be used in more effective ways to improve student learning.

The Brookings Papers on Education Policy, edited by Diane Ravitch, and published in 2005, examined both sides of the debate over high-stakes standardized testing. Tom Loveless, in his essay, claimed that the recent education reform, under Presidents Clinton and Bush, calling for standards and standardized testing, have indeed provided important incentives for educators to improve student learning. The studies he cited are mixed, because under NCLB, each state has been allowed to design their own standards and standardized testing mechanisms. There is no confirmed causal effect for why some

---

<sup>78</sup> Thomas J. Kane and Douglas O. Staiger, “The Promise and Pitfalls of Using Imprecise School Accountability Measures,” *The Journal of Economic Perspectives* 16, no. 4 (Autumn, 2002). <http://www.jstor.org/stable/3216916>. (accessed October 15, 2010), 92.



states or groups perform better than others. Despite the mixed results, he saw a greater push for academic achievement in the subjects tested like English and math. Across the United States, he considered the results a positive step toward achieving needed educational reform.<sup>79</sup> However, he did not understand why teachers, students, and parents have formed coalitions in multiple states to push against the high-stakes standardized testing. He claimed, “Accountability targets marginal producers. Teachers and students who fall into that category--or who fear falling into that category--are a ready-made group of opponents.”<sup>80</sup> Throughout his essay he identified opponents to the high-stakes standardized testing movement as progressives who favor “real world” learning and “authentic assessment.” Without explicitly stating it, he set up an *if/then* argument that *if* someone opposes standardized testing, *then* they must be “progressive” and “marginal producers.”<sup>81</sup>

The essay written by Tom Loveless was followed with comments by two education policymakers: Robert M. Costrell, who supported the findings from Loveless, and Larry Cuban, who questioned the overall goals of standardized testing. Costrell and Loveless shared the same enthusiasm for high-stakes standardized testing, but they both used only one type of high-stakes standardized test to support their reasoning, mandatory high school exit exams. The accuracy results of high school exit exams is much higher than NCLB mandated standardized tests. The question of accuracy is one of the major reasons why opponents of standardized tests criticize the results. High-stakes standardized testing under NCLB are only high-stakes for teachers, administrators, and schools, not for

---

<sup>79</sup> Tom Loveless, “Test-Based Accountability: The Promise and the Perils,” in *Brookings Papers on Education Policy*, ed. Diane Ravitch (Washington D.C.: Brookings Institution Press, 2005), 8-14.

<sup>80</sup> Loveless, 25.

<sup>81</sup> Loveless, 21-25.

individual students. In reality, these tests under NCLB have no real-world consequence for the students besides a label of “Far Below Basic,” to “Advanced,” which sometimes, depending on the school, will decide remedial placement, or advanced placement.

Outside of placement in a class and a personal stigma, students have no real incentive to do their best. Why would students who are not motivated by grades or scores try their best on a 300-500 question multiple choice test that takes between four to eight days to complete? Why would students that otherwise perform well in school, try their best on a test that doesn't really personally matter to them? Eleventh graders are the worst type of tester for NCLB related tests, because the students at that point really do not have an incentive to try their best. If they are college bound, then they have already completed the SAT, just months before, and are in the process of applying to college. Eleventh grade students generally do not care about placement in their senior year courses, because they have either already met all their college admissions requirements, or they are taking basic level courses anyway, so they have nothing to lose. It is illogical for juniors in high school on both ends of the spectrum between “far below basic” and “advanced” to try their best on the state standardized tests mandated by NCLB.

In contrast, the accuracy of high school exit exams is much higher, because if the students don't pass the test, then they don't receive a diploma, even if they have met all other graduation requirements. Most of Costrell's analysis focused specifically on his home state of Massachusetts, where the failure rate of the exit exam was reduced by 20 percentage points the first year the test mattered. Costrell claimed, “...it took a certain faith in our students' potential and our teachers' skills, as well as in the logic of human behavior, to believe that the failure rate would drop as dramatically as it did once the test

mattered.”<sup>82</sup> Costrell’s reasoning could be used to argue against the accuracy of standardized tests, because the dramatic increase of 20 percentage points in one year (once the test mattered) could speak more to the fact that students actually took the test seriously, rather than instruction or learning suddenly improving. His findings raised the question of the accuracy of standardized test scores that have no direct bearing on individual students.

Costrell concluded his comments with the notion that finding a balance between intervention and flexibility is difficult but necessary. He posed the question, “...should we provide underperforming districts with greater flexibility (by, for example, suspending collective bargaining restrictions on personnel deployment) in the hope that such flexibility will be well used, or should we actively intervene, taking over the school and telling it what to do?”<sup>83</sup> Costrell hoped that policymakers will stay the course and allow standards-based accountability to continue so that research will progress far enough to determine effective solutions. Loveless and Costrell both advocated for further research to show how to raise test scores rather than research to search for authentic assessment tools that align closer with skills needed today.

Larry Cuban challenged the long term goals of supporters of standardized testing, specifically the goals defined by Loveless and Costrell. He challenged the assertions by Loveless that coercive accountability of students is positive when the testing movement has produced mixed results that do not conclusively define good teaching, or how

---

<sup>82</sup> Loveless, 32. These comments are by Robert M. Costrell in a follow-up to the article by Loveless.

<sup>83</sup> Loveless, 36. These comments were written by Robert M. Costrell in a follow-up to the article by Loveless. The collective bargaining issue that Costrell brings up is a common criticism from supporters of standardized testing. Some opponents of standardized testing argue that the “take-over” measure designed in NCLB is a tool to dismantle teachers’ unions.

effective learning occurs. Cuban pointed out that certain assumptions continue to drive the coercive testing movement.<sup>84</sup> A driving force behind these assumptions is the idea that students must be able to perform well in academic subjects such as math, science, and English, when in reality only a certain percentage of our society needs to be advanced in these academic areas. He critically asked, “Will we be a nation of lawyers, engineers, managers, and teachers? Who will wait tables, fix broken pipes, assemble computers, sell products, and empty bedpans?” Out of the students that manage to enter college, fewer than 60 percent finish with a four year degree. The rate of failure for minorities is even higher.<sup>85</sup> Cuban reminded his readers of the many other important skills in life that are not given enough priority in an educational system driven by high-stakes standardized testing.

A few years after Cuban’s debate with Loveless and Costrell, Harvard published a media study that provided evidence that supported Cuban’s ideas that high-stakes standardized testing creates an educational system that focuses its energy on a limited group of students. This book exposed the trickery used by educators to manipulate test scores, so the idea of not leaving any students behind (*No Child Left Behind*), has led to

---

<sup>84</sup> Loveless, 39-40. These comments were written by Larry Cuban in a follow-up to the article by Loveless.

The assumptions that Cuban refers to are: 1. Strong economic growth, high productivity, long-term prosperity, including a higher standard of living, and increased global competitiveness depend upon a highly skilled workforce. 2. Public schools are responsible for equipping students with the necessary knowledge and skills to compete in an information-based workplace. 3. All Public schools are doing a poor job of preparing high school graduates for college and the workplace, with urban schools doing the worst job of all. 4. Schools are just like businesses. The principles that have made businesses successful can be applied to schools to produce structural changes that will improve academic achievement as measured by standardized tests, end skills mismatch, and increase public confidence in schools. 5. Higher test scores in school mean future employees will perform better in college and in the workplace, 39-40.

<sup>85</sup> Loveless, 40. These comments are by Larry Cuban in a follow-up to the article by Loveless.

efforts by schools to leave some students behind.<sup>86</sup> The authors studied hundreds of media reports regarding allegations of cheating by teachers or administrators on standardized testing results. They wrote, “Our goal was to present a cohesive and convincing set of examples of the problem associated with high-stakes testing. We hope this will convince legislators and other supporters of high-stakes testing that the costs associated with high-stakes testing are simply not worth it.”<sup>87</sup>

One of the tools they used to make their argument is a social science theory known as the principle of Campbell’s law. When used with high-stakes standardized testing, this law theorizes that the opposite effect desired by NCLB legislation occurs with testing, because “the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it was intended to monitor.”<sup>88</sup> Instead of holding schools accountable to see how they serve their lowest students, they merely pushing the lowest students out so their scores won’t appear in the testing data. For example, a Tennessee newspaper reported, in a survey of teachers and administrators, that nine percent of the teachers surveyed said they had witnessed test impropriety on the state standardized tests; this is at the same time Tennessee is connecting standardized test results to teacher merit pay.<sup>89</sup> Some of the “adult cheating” they noticed was removing academically lower

---

<sup>86</sup> For another example of teachers cheating, see: Steven D. Levitt and Stephen J. Dubner, *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything* (New York, Harper Perennial, 2009), 222-223.

<sup>87</sup> Sharon L. Nichols and David C. Berliner, *Collateral Damage: How High-Stakes Testing Corrupts America’s Schools* (Cambridge: Harvard Education Press, 2007), xviii.

<sup>88</sup> Nichols, 26-27.

<sup>89</sup> The temptation to cheat is even greater when the results are connected to merit pay. Tennessee is of particular interest for this paper, because it is the first state that adopted a teacher “performance” pay system similar to the TAP program recently adopted by Sycamore River--the next section will document the Tennessee Value-Added Assessment System (TVAAS) in more detail.

performing students from testing by putting them in the library for a week, or creatively finding excuses to suspend them from school during testing week.<sup>90</sup> They categorized other forms of cheating they noticed into three major areas: first, pretest cheating activity, which involved teaching to the test by using the tests or similar tests to prepare students for the test questions; second, during-the-test cheating activity, which spanned incidences of coaching the students to help them arrive at the correct answer to actually telling them the correct answers to bubble in; third, post-test cheating activity, which involved going back and changing students bubbled answers on standardized tests, or in a New York case, “scrubbing” English essay exams by a third teacher brought in to subjectively find extra points above what the first two teachers, who originally scored the exam, found in hopes that the student would then have enough points to pass the exam.<sup>91</sup> This shows that under a high-stakes testing environment, the incentive to cheat exists, even on more authentic assessments like essay exams.

One way to make cheating more difficult is to use multiple measures of assessment. Multiple measures not only make cheating more difficult, they also ensure a more accurate understanding of what the student knows and is able to do. In 2009, Susan Brookhart published an article in *Educational Leadership* reminding educators of several guidelines defining ethical conduct and legal obligations of educators regarding multiple measures of assessment. The *Code of Professional Responsibilities in Educational Measurement* (Section 6,7), from the National Council on Measurement in Education (1995), stated, “Persons who interpret, use, and communicate assessment results have a professional responsibility to use multiple sources and types of relevant information

---

<sup>90</sup> Nichols, 36.

<sup>91</sup> Nichols, 37-45.

about persons or programs whenever possible in making educational decisions.”<sup>92</sup> The same principle found in *The Standards for Educational and Psychological Testing* stated, “In educational settings, a decision or characterization that will have major impact on a student should not be made on the basis of a single test score. Other relevant information should be taken into account if it will enhance the overall validity of the decision.”<sup>93</sup> Brookhart explained that these standards are no longer just principles. In 1994, in Title I of the *Improving America’s Schools Act*, similar language required educators to use multiple measures of assessment to judge the performance of schools. In 2001, NCLB adopted the same language from the *Improving America’s Schools Act*, “Such assessments shall involve multiple up-to-date measures of student academic achievement, including measures that assess higher-order-thinking skills and understanding.”<sup>94</sup>

Educational experts designed the guidelines and the legislators adopted them. Unfortunately, somewhere along the way the spirit of the law that clearly required multiple measures of assessment to truly understand what a student knows and is able to do got lost in the standardized testing movement. Brookhart explained that in 2004, NCLB guidelines qualified “multiple measures” being met by any assessments that measured higher-order thinking, so some states claimed they met the obligation by giving multiple opportunities to pass the same type of high school exit exam.<sup>95</sup>

Brookhart’s argument reflects current thinking about how standardized test scores are inadequate indicators of student achievement. In the same issue of *Educational*

---

<sup>92</sup> Susan M. Brookhart, “The Many Meanings of ‘Multiple Measures,’” *Educational Leadership* 67, no. 3 (November 2009), 8.

<sup>93</sup> Brookhart, 8. *The Standards for Educational and Psychological Testing* come from the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999. The standard listed here is Standard 13.7.

<sup>94</sup> Brookhart, 8.

<sup>95</sup> Brookhart, 8.

*Leadership*, Gerald Bracey also challenged the results of standardized tests. According to Bracey, in our current climate of NCLB and teacher merit pay, our society is pushing education for better results. Instead of creating long term improvements in education, the emphasis on standardized testing is creating a system that implements measures to produce short term improvements in test scores. It is really not about the child or the instruction. Everything hinges off the test scores for the districts or schools that are targeted as program improvement schools. Bracey explained the problem with such emphasis on standardized testing by exposing some fallacies around testing.<sup>96</sup> One fallacy is that the United States is falling behind other nations. When comparing tests from the United States with other countries, several problems arise like language and cultural differences found within homogeneous populations in some nations like Singapore versus diverse populations as found in the United States. For example, one test result that caused alarm in the United States is the fact that only about one-third of eighth graders read at a “Proficient” level. Bracey used the findings of one study to show that, “If students in other nations took the NAEP, only about one-third of them would also score ‘Proficient’--even in the nations scoring highest on international reading comparisons.”<sup>97</sup>

Another fallacy is the idea that a nation’s economy is dependent on high standardized testing results. Bracey poses a question that challenges this fallacy. Why

---

<sup>96</sup> Gerald Bracey, “Big Tests: What Ends Do They Serve,” *Educational Leadership* 67, no. 3 (November 2009): 32-37.

<sup>97</sup> Bracey, 33. The study he refers to was published by Richard Rothstein, Rebecca Jacobsen, and Tamara Wilder (2006, November 29). Proficiency for all is an oxymoron. *Education Week*, 26 (13), 32,44.



have the economies of Japan and Iceland collapsed in recent decades when both have led the world in performance on standardized tests?<sup>98</sup>

Another fallacy is the idea that performing well on standardized tests means a person possesses what he or she needs to be successful. Bracey argued against this fallacy by explaining that most standardized testing does not measure creativity, curiosity, or a sense of adventure, three characteristics that have historically given the United States an advantage regarding innovation. Singapore's Minister of Education, Tharman Shanmugaratnam, found that students in Singapore score well on standardized tests but, 10-20 years later, are not as successful as students in the United States. Bracey and Shanmugaratnam credited this notion of success to the fact that the education system in the United States still values creativity and innovation. However, with increased emphasis on standardized testing the United States is slowly losing the creativity and innovation practiced in school. One researcher concluded, "The increasingly massive and far-reaching use of standardized tests is one of the most effective, if unintentional, vehicles this country has created for suppressing creativity."<sup>99</sup>

At a time when creativity is vital for the competitive innovation needed by the United States, there exists a huge disconnect between the educational research that finds standardized tests inadequate for accurately measuring authentic student achievement and education policy that continues to push standardized testing. Parents, voters, and even many teachers seem blind to the issues surrounding standardized testing, and the growing chasm between policy and research results. The public wants to see results, but they rely on education to assess the learning that shows that their children are prepared for the

---

<sup>98</sup> Bracey, 35.

<sup>99</sup> Bracey, 35. The researcher quoted here is Robert Sternberg in his article "Creativity is a habit." *Education Week*, (2006, February 22), 47.

future. The media stirs society for educational reform without educating their audience about the well established fact that standardized tests are not the answer. The debate in the early 1990s about authentic measurements involving portfolios and project-based learning seems to fade as more time passes. The NCLB crunch-year of 2014, when *one hundred percent of students will be “Proficient”* looms on the horizon. The ignorance of educational research courts the latest craze for educational reform, teacher merit pay, a perceived snake-oil, miracle-tonic reform that has cycled through educational reform history.

It is time for policymakers to turn away from the status quo; it consistently proves to be ineffective. Instead of continuing the nonsensical obsession with standardized testing, policymakers should turn where researchers mapped, where brave teachers and schools trod successfully, where students grow equipped for critical thinking and independent learning. Unfortunately, teacher merit pay is an idea that has failed many times before, but is again appealing to policymakers who seem blind to history or educational research. This time around teacher merit pay is equipped with a new gimmick. Beginning in the 1990s, teacher merit pay systems began using the results from standardized testing to evaluate teachers. Now, the temptation is even greater for education policy to continue with the habit of standardized testing and miss an opportunity for lasting reform. Why do teachers, for the most part, remain quiet on these issues, besides an occasional editorial or an official union statement or action? Out of the millions of teachers in this country, the number that stand up against standardized testing is slim to none. Because it is their profession, teachers possess a special responsibility to know what the research finds. After considering the research, certainly the first step for

teachers is to collaborate how to resist any further emphasis on standardized testing, an emphasis that is currently happening with teacher incentive programs like TAP.

Communities need to hear from their teachers why standardized testing is not accurate and authentic and should be dissolved rather than receiving a promotion through teacher performance pay systems.

Why is the resistance from teachers not happening at this point? Maybe, even the teachers that know what the experts know, experience the same phenomena as Americans with their fast food obsession. Many people across America understand how unhealthy fast food is, and yet they continue to consume it, because it is cheap, fast, easy, and familiar. It also tastes good. Maybe the lack of resistance by teachers has to do with the added seasoning of teacher merit pay that has just made standardized tests more palatable for those teachers receiving a bonus check. However, as the experts warn, just because it tastes good, doesn't mean it's good for you.

## Chapter IV

### The Pushing and Resisting of Teacher Merit Pay<sup>100</sup>

The implementation of teacher merit pay in the past attempted to bring educational reform, especially reform in the classrooms that housed the poorest children. Even hundreds of years ago reformers were concerned that some children were being left behind. With the right intentions, reformers believed that teacher merit pay would ensure that all children receive a quality education. If there exists a lack of understanding of what truly occurs in a classroom, and if there somehow exists a notion that classrooms are like production lines, then it is reasonable to think that student achievement will go up if teachers are paid to work harder and better. This type of rhetoric possesses coded business language that is driven by multiple assumptions. First, is the assumption that society (the consumer) can agree upon the outcome of what the student (product) should know and be able to do. Second, is the assumption that when the students are in the classroom (all the raw materials), then the learning (factory production) can occur under the full control of the teachers (producers). Third, is the assumption that the only valuable learning (production) is the student achievement (finished product) that can be measured in a standardized (quality control) way.<sup>101</sup>

---

<sup>100</sup> TAP does not use the term merit pay. That system uses performance pay. As far as my research shows there is no real difference between the two terms. Systems similar to TAP have used the term merit pay long before TAP ever came along. This paper will use merit pay and performance pay interchangeably.

<sup>101</sup> Henry C. Johnson, ed. *Merit, Money and Teachers' Careers: Studies on Merit Pay and Career Ladders for Teachers*. Lanham: University Press of America, Inc., 1985. This book is a collection of essays that brings researchers together who contemplate how a "performance-based pay" system should operate within education. Many of the researchers use business as the comparative model.

Another book proposes a similar corporate philosophy for educational reform. The fix, according to the corporate leaders, is to pay teachers performance pay so they will work harder and smarter. This book is: Richard M. Brandt, *Incentive Pay and Career Ladders for Today's Teachers: A Study of Current Programs and Practices* (Albany: State University of New York Press, 1990).

When considering how corporate reformers try to equate education with business, Diane Ravitch wrote, “[They] betray their weak comprehension of education...They think they can fix education by applying the principles of business, organization, management, law, and marketing and by developing a good data-collection system that provides the information necessary to incentivize the workforce.”<sup>102</sup> The multiple assumptions, caused from thinking that education works like business, miss the fact that humans are the most dynamic beings on the face of the earth, possessing a legion of forces outside the control of the teacher. They change everyday, sometimes moment by moment, their hopes, their attention span, their retention ability, all because of who they are, where they came from, and where they are going. Did Johnny’s dad hit his mom last night? Did Juan get any food to eat this morning? Did Jessica swallow multiple pain medications? Will Jun Li even show up today? If these students were workers, many of them would lose their jobs. The employer wouldn’t put up with their behavior or their lack of attention, but teachers do not have the same luxury of selecting the workers who sit in the classroom. Society must consider these difficult concerns to stop thinking that teachers need to just work harder and everything will be okay.

Of course teachers have a responsibility to do everything within their power to wade through the muck tramped into the classroom on the souls of their students’ lives, but to think that merit pay and standardized tests are the solution to our education woes is to miss the fact that our students are individuals with hopes and dreams, scars and disappointments, and curiosity with an ability to solve complex problems. The idea that teachers are the producers completely misses an understanding of how learning should

---

<sup>102</sup> Diane Ravitch, *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education* (New York: Basic Books, 2010), 11.

occur. Teachers facilitate learning. The students are the producers that should be pictured in our collective mind standing over their creation, their innovation, their future.

Why are students pushed down into a chair and given only four multiple choices?

The attempt to fix education with the notion that money will make harder working teachers, that will produce a better product, that can be measured in a standardized way, spanning centuries has failed eventually, every time it has been attempted.<sup>103</sup> Teacher merit pay systems fail eventually for several reasons: cheating, curriculum narrowing, and the realization that the standardized test results used to evaluate teachers are not accurate or authentic. Even if one places value in the results of standardized tests, a three-year study revealed, “Overall we find no effect of teacher incentives on student achievement.”<sup>104</sup> A survey of studies regarding teacher performance pay and increased student achievement concluded, “...considered together these studies failed to support any clear inferences regarding achievement effects associated with performance pay programs.”<sup>105</sup>

What is the purpose of teacher evaluations? If it is to encourage better teaching, then is there a way to introduce more incentive into education so that teachers are

---

<sup>103</sup> Terry Herndon, “Merit Pay and the Concerns of the Teaching Profession,” in *Merit, Money and Teachers’ Careers: Studies on Merit Pay and Career Ladders for Teachers*, ed. Henry C. Johnson, Jr. (Lanham: University Press of America, Inc., 1985). In this essay, Herndon points out that society calls for merit pay, but they really don’t understand how merit pay adversely affects education. Even when it has a failing track record, society still pushes it, because the next generation doesn’t realize the merit pay failed previously.

<sup>104</sup> Matthew G. Springer, Dale Ballou, Laura Hamilton, Vi-Nhuan Le, J.R. Lockwood, Daniel F. McCaffrey, Matthew Pepper, and Brian M. Stecher. “Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching.” National Center on Performance Incentives, Project on Incentives in Teaching (September 21, 2010). That statement found on page 36 of the report. <http://edlab.tc.columbia.edu/files/pointstudy.pdf> (accessed October 22, 2010).

<sup>105</sup> Edward W. Wiley, Eleanor R. Spindler, and Amy N. Subert. *Denver Pro Comp: An Outcomes Evaluation of Denver’s Alternative Teacher Compensation System-2010 Report*. That statement found on page 15 of the report. <http://static.dpsk12.org/gemsnewprocompProCompOutcomesEvaluationApril2010final.pdf> (accessed 10/20/2010).

rewarded when they are actively trying to improve their teaching? Performance pay must be an encouragement to improve skills as a teacher rather than try to improve test scores, which the teacher doesn't have complete control over. In 1990, a group of researchers analyzed different forms of merit pay systems to see if they truly encouraged quality teaching. They studied various forms of evaluation systems used for teacher merit pay, some of which stood in direct contrast to each other like: skill-based rather than performance-based evaluation; developmental rather than uniform evaluation criteria; subjective rather than objective evaluations; and formative rather than summative evaluations.<sup>106</sup> According to these researchers, there are two ways to approach merit pay for teachers: the *old style* merit pay, which assesses teacher performance in the classroom focusing on presentation style and organization of the content of lessons, and the *new style* merit pay, which rewards teachers based on output measures like student performance on standardized tests.<sup>107</sup> Both kinds of merit pay have problems. The problems with the New Style merit pay are still being explored, because it didn't become a practice in schools in the United States until the 1990s when the motivation behind teacher merit pay systems changed, but nearly two decades of research do show severe problems with any form of teacher merit pay that is connected to standardized test results. That research will be explored in this section.

A handbook written in 1984 explained a different motivation behind teacher merit than the current motivation to use performance-based pay as a way to hold teachers

---

<sup>106</sup> Samuel B. Bacharach, Sharon C. Conley, and Joseph B. Shedd, "Evaluating Teachers for Career Awards and Merit Pay," in *The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers*, ed. Jason Millman and Linda Darling-Hammond (Newbury Park: Sage Publications, Inc., 1990), 1-39.

<sup>107</sup> Bacharach, "Evaluating Teachers for Career Awards and Merit Pay," 26-28.

accountable.<sup>108</sup> This handbook was published by three different educational administration associations with a different motivation than exists today, to give teachers access to higher pay, not to look for ways to penalize teachers whose students don't perform well on standardized test scores. The handbook suggests that many forms of measurement should be used, and like so many other studies conclude, it warns that test scores are the result of many things outside the control of the teacher.

The following year, in 1985, the National Association of Secondary School Principals weighed in on the various issues surrounding teacher merit pay. One study cited in the booklet looked at fifty years of incentive pay, and concluded “pay for performance leads to an increase in performance.”<sup>109</sup> With that evidence, the study asked a critical question--at what cost did that increase in performance come. The challenge educational reformers face today is to measure performance connected to the desired outcome. If performance pay is largely based off standardized test scores, then eventually we should see an increase in test scores, pulling student learning away from the higher order thinking skills required in the 21st century.

According to James C. Parker, in order for incentive pay to work, one of several criteria that must be met is, “The individual (teacher) must see a clear relationship between the reward and the required task.”<sup>110</sup> This observation directly relates with the current debate over teacher merit pay that uses value-added methodology. Placing too much emphasis on standardized test scores won't work, because the required task that is

---

<sup>108</sup> McCormick Cresap and Paget Cresap, *Teacher Incentives: A Tool for Effective Management* (Reston: National Association of Secondary School Principals, National Association of Elementary School Principals, and American Association of School Administrators, 1984).

<sup>109</sup> James C. Parker, *Career Ladder/Master Teacher Programs: Implications for Principals* (Reston: The National Association of Secondary School Principals, 1985), 13.

<sup>110</sup> Parker, 14.



connected to the reward (the performance of students on standardized tests) is not essentially within the control of the teacher. Value-Added Methodology claims that it gets away from this problem, because it doesn't take just one snapshot-test score and compare it to other students. Instead, it compares last year's test score with this year's test score and measures the individual growth of each student. For example, if a student scored "Far Below Basic" last year, and now this year they scored a "Below Basic," then that teacher is credited for increasing the achievement growth of that student, even though the student is still considered low.<sup>111</sup>

The idea that *two* snapshot-test scores provide a more accurate picture of the student achievement results is simply missing the reality of the struggles that so many students deal with outside of school. These two snapshot comparisons still do not take into account outside forces that affect student achievement. What if the student's parents were doing fine last year, but now they are separated and going through a divorce? What if the student didn't have a drug problem last year, but now he does? What if the student has the flu during the test or during the time the lesson was taught and practiced in class? What if the student forgot to take his ADHD medication that day that math was tested, or what if during the several days that the math concept was taught his medication level was not the correct level? What if the single mom of the student lost her job, and now the student is living in a more crowded chaotic environment? The snapshot assessment administered once a year by a limited multiple choice test, comparison or no comparison, is what makes the value-added methodology so weak. Two bad measurements compared together does not equal an accurate picture of student achievement. Instead, to obtain a

---

<sup>111</sup> Evergreen Freedom Foundation, "Value Added Assessment," *Evergreen Freedom Foundation, School Director's Handbook* (n.d.) <http://www.myfreedomfoundation.com/pdfs/Value-Added.pdf> (accessed October 25, 2010), 3.

more accurate understanding of student learning and to attempt the best effort to separate outside forces in students' lives from teacher performance, multiple assessments need to be administered once or twice a month, and many of those assessments need to be authentic in nature where students have more flexibility to demonstrate competency without the limitations of multiple choice-bubble tests. Later, this section will unfold more research findings regarding the weakness of standardized test--value-added methodology.

Merit pay keeps cycling through educational reform history. Why? People support merit pay for a variety of reasons. One major reason given repeatedly by supporters is that money motivates, so “teachers will raise their performance if offered a pay raise for doing so.”<sup>112</sup> Another argument made by supporters of merit pay claims that merit pay will provide a better system of feedback to teachers rather than the existing system that allows teachers to define their own standards of effectiveness.<sup>113</sup> Another argument in favor of merit pay is that it provides more power to administrators to achieve staff coordination, introduces new methods and aims, and places administrators in more control over the process of teaching.<sup>114</sup> Retention and recruitment of quality teachers increases, according to supporters of merit pay, because these quality teachers are rewarded for their efforts, and the public is more willing to increase spending on education if they know the increase in spending goes directly to quality teachers.<sup>115</sup>

What prevents merit pay from being successful long term? One concern raised by opponents of merit pay asserts that the content of teaching becomes highly directive to

---

<sup>112</sup> Bacharach, “Evaluating Teachers for Career Awards and Merit Pay,” 5.

<sup>113</sup> Bacharach, “Evaluating Teachers for Career Awards and Merit Pay,” 6-7.

<sup>114</sup> Bacharach, “Evaluating Teachers for Career Awards and Merit Pay,” 8.

<sup>115</sup> Bacharach, “Evaluating Teachers for Career Awards and Merit Pay,” 10-12.

meet the criteria used to judge performance. This negatively affects the relationship between teachers and students and the intrinsic motivation of teachers. The “intensely satisfying experience of educating minds” is a critical resource and diminishes with merit pay, opponents argue.<sup>116</sup> Another argument against merit pay is that “average” teachers become discouraged and experience an incentive to lower their performance. Similar to this argument is the result of many studies that show that merit pay systems “foster dissension, rivalry, and jealousy among teachers.”<sup>117</sup> Other concerns regarding merit pay, according to the studies analyzed are: the distrust that affects teacher-administrator relationships; the impact of parent-school relations with parents advocating to move their children into classrooms where teachers received merit pay; judgmental versus diagnostic feedback where the blame-game smothers the effort to explore ways to improve; and the impact on staff development, because merit pay can result in competition among teachers rather than cooperation.<sup>118</sup>

The difficulty of measuring performance is another major obstacle that prevents long-term success of merit pay systems. The researchers clearly stated, “Even if the motivation arguments for merit pay are supportable, they are irrelevant if merit cannot be measured validly and reliably. An evaluation system is valid if it measures what it is supposed to be measuring, and is reliable if the measurement instrument produces consistent measures.”<sup>119</sup> This is exactly why any merit pay system that relies heavily on standardized test scores will not work. Outside influences and one testing sample per

---

<sup>116</sup> Bacharach, “Evaluating Teachers for Career Awards and Merit Pay,” 16-19.

<sup>117</sup> Bacharach, “Evaluating Teachers for Career Awards and Merit Pay,” 20.

<sup>118</sup> Bacharach, “Evaluating Teachers for Career Awards and Merit Pay,” 23-26.

<sup>119</sup> Bacharach, “Evaluating Teachers for Career Awards and Merit Pay,” 26.

year both affects the reliability of the test results, and the multiple choice structure of the test severely hinders the validity of the results.

The 1990 research study concluded with the idea that the theoretical advantages and disadvantages of teacher merit pay could be argued indefinitely, but “the acid test lies in its implementation,” and findings of the researchers show a discouraging historical experience. Their case studies from the 1930s, 1950s, and 1970s showed that the majority of teacher merit pay systems fail within a few years mainly because of unsatisfactory evaluation systems and teacher dissension caused from teachers feeling the merit pay system was unfair, created distrust, resentment, and conflict between teachers and teachers and administrators.<sup>120</sup> This historically repeated failure of teacher merit pay and the fact that standardized test scores are not accurate or authentic should raise questions about any system that connects performance pay with standardized test results.

Even the groups that traditionally support standardized testing continue to have reservations about connecting standardized testing results to teacher accountability. In 2005, the National Research Council (NRC) and the National Academy of Education published a report from a workshop funded by Carnegie Corporation to ascertain whether the current push for Value-Added Assessment (VAA) will result in improved student learning. Value-Added Assessment is re-emphasizing standardized test results at a time when educators attempt to de-emphasize standardized testing and focus more exclusively on project-based learning and the skills recognized as 21st century learning skills. The teacher incentive program that drives Sycamore River’s implementation of its TIF (Teacher Incentive Fund) is the TAP system of teacher evaluation. The TAP system uses

---

<sup>120</sup> Bacharach, “Evaluating Teachers for Career Awards and Merit Pay,” 37.

standardized testing to determine half of the teacher evaluation rating.<sup>121</sup> Computation of the standardized test results runs through a process known as Value-Added Assessment. The impetus behind the workshop was to identify areas of strength and areas of weakness behind the various components of VAA: school and teacher improvement, program evaluation, and research. One of the major observations from the report was that tests are incomplete measures of achievement. The report focused specifically on value-added methods of teacher-pay-for-performance (the same way TAP is designed for Sycamore River), and warns of the challenges of using value-added methods:

Value-added methods involve complex statistical models applied to test data of varying quality. Accordingly, there are many technical challenges to ascertaining the degree to which the output of these models provides the desired estimates. Despite a substantial amount of research over the last decade and a half, overcoming these challenges has proven to be very difficult, and many questions remain unanswered--at a time when there is strong interest in implementing value-added models in a variety of settings.<sup>122</sup>

The report admitted that test scores and students' rate of growth in achievement are both highly correlated to students' socioeconomic status (SES).<sup>123</sup> Economic status is one of the most powerful outside forces affecting student achievement. The workshop participants expressed concerns about implementing VAA for several major reasons. Uses and possible consequences of VAA created one major area of concern. By using value-added models, tested content will be prioritized over non-tested content, which brings up concerns of other researchers that the teaching of 21st century learning skills and project-based learning will suffer. Another concern stated, "The evidence for the

---

<sup>121</sup> The TAP system will be explained more fully in the next section.

<sup>122</sup> Henry Braun, Naomi Chudowsky, and Judith Koenig, eds., *Getting Value Out of Value-Added: A Report of a Workshop, Committee on Value-Added Methodology for Instructional Improvement, Program Evaluation, and Educational Accountability* (Washington, D.C.: The National Academies Press, 2010), vii.

<sup>123</sup> Braun, 6.

reliability and validity of value-added estimates is not sufficiently strong to support their use as the sole basis for high-stakes decisions, and therefore they are most appropriately used in combination with other indicators for such purposes.” For this reason, the TAP system of teacher performance pay weighs standardized test results as fifty percent of its formula. Under the TAP system, there is still room for significant statistical error when considering that the half of the data (formed from standardized test results) could be inaccurate for a variety of the reasons that are raised by the workshop participants. Another concern from workshop participants dealt with the question of attribution like the question of how individual teachers can be evaluated in a case of team teaching.<sup>124</sup>

Measurement issues formed another major area of concern from workshop participants. The group of participants recognized that: tests have limitations for assessing all the important elements of student achievement; results are not precise; interval scales do not provide consistent rankings of schools, teachers, and programs; vertical linking of tests is difficult to attain, because content from two grade levels do not compare; and developmental levels of learning are not built into the comparison formula.<sup>125</sup>

The last major area of concern from workshop participants dealt with analytic issues. Bias is built into the model for analyzing VAA, so the concern is the overestimating and underestimating of previous school or program effects. Precision and stability raise concern, because of the errors that are consistently found in samples. VAA incentive pay for individual teachers becomes controversial over the idea that a few low performing students can pull a teacher’s performance rating down, and conversely a few

---

<sup>124</sup> Braun, 12-13. This reference is for all material cited in this paragraph.

<sup>125</sup> Braun, 12-13.

high performing students can raise it. In reality, a good teacher could get a bad rating and a bad teacher could get a good rating, all depending on the extreme results of a few students. Data quality raises another concern, because the value-added estimates rely so heavily on data that any missing or faulty data can throw the results off, especially when a layer of variables form the results. Complexity versus transparency is the last analytic concern of participants. Balancing the two creates a tradeoff between more accurate results created from added complexity or keeping the level of complexity down so that the results can be understood by educators and parents.<sup>126</sup>

An article that appeared in *Educational Leadership* (2005) shared the same concerns as the workshop participants. In the article, Harold Doran and Steve Fleischman argued that Value-Added Assessment is not an accurate tool to measure teacher effects, because “as it currently stands, no empirical research validates the claim that value-added models accurately identify the most effective teachers. The many anecdotal claims have not yet been verified through experimental research.”<sup>127</sup> According to the authors, one of the many reasons why value-added assessment does not work is that VAA designers assume a student’s growth can be measured from year to year, but that assumption does not take into account the fact that sometimes curriculum is not similar and therefore impossible to compare year to year. This situation of non-comparable curriculum occurs more often at the secondary level. For example, health is taught in the ninth grade and not followed by any comparable curriculum the next year. It also fails to recognize mitigating circumstances in the student’s personal life that may affect performance on the test. These circumstances change year to year. Doran and

---

<sup>126</sup> Braun, 12-13.

<sup>127</sup> Harold C. Doran and Steven Fleischman, “Challenges of Value-Added Assessment,” *Educational Leadership* 63, no. 3 (November 2005), 86.

Fleischman list several other studies with findings that showed that an introduction of vertical scales into VAA statistical analysis can introduce more error in longitudinal analyses.<sup>128</sup>

One place where the research findings from Doran and Fleischman were observed was Tennessee. In 2004, Lynn Olson wrote a short article that discussed the results of Value-Added Assessment in Tennessee. Her article gave a key explanation to why VAA is controversial. The article reported, “Much of the frustration does not center on the value-added method itself, but on a step its creator, William Sanders, takes before calculating the scores. Essentially, he tries to determine if the scale used to weigh the easiness or difficulty of individual test items is equivalent from year to year, so that test results are comparable from one year to the next.”<sup>129</sup> This weighted scale can create discrepancies between the test scores and the VAA results, as much as 40 percent or more, educators in Tennessee realized. Supporters of TAP argue that only half of the teacher evaluation is based off standardized test results, so they would argue that critics should relax their concern over standardized test results connected to the teacher evaluation. However, this is a weak argument when considering that VAA results can create such a large discrepancy between the outcome of the formula and the actual test

---

<sup>128</sup> One of those studies is: Doran, H.C., and Cohen, J. (2005) The confounding effect of linking bias on gains estimated from value-added models. In R. Lissitz (Ed.), *Value-added model in education: Theory and applications*. Maple Grove, MN: JAM Press. The other study is: Michaelides, M.P., and Haertel, E.H. (2004, May) *Sampling of common items: An unrecognized source of error in test equating* (Technical Report). Los Angeles Center for the Study of Evaluation and National Center for Research on Evaluation, Standards, and Student Testing.

This problem of non-comparable curriculum and the fact that many secondary students do not perform with their greatest efforts on state standardized tests are probably the two main reasons why TAP systems experience implementation difficulties at the secondary level. This fact was established by a TAP representative during a question and answer exchange during a presentation at a middle school in front of a room full of Sycamore River teachers on November 12, 2010

<sup>129</sup> Lynn Olson, “Tennessee Reconsiders Value-Added Assessment System,” *Education Week* 23, no. 25: 9 (2004) *Academic Search Elite*, EBSCOhost (accessed October 26, 2010), 2.



results. A somber reminder of the impact of half of an evaluation occurs when remembering how much failing a college course final exam (worth half the class) can negatively affect the overall percentage of the class. The article reported, “Concerns mounted starting in the fall of 2002, when the 71,000-student Nashville public schools noted significant discrepancies between what the raw test results showed and what the value-added measures showed, after Mr. Sanders had made his adjustments, particularly in fourth grade reading and language arts.”<sup>130</sup>

In the preliminary report of a three-year study (previously mentioned in this section) of the Metro-Nashville Public Schools from 2006-07 through 2008-09, a conclusive finding was powerfully summarized, “Overall we find no effect of teacher incentives on student achievement.”<sup>131</sup> This study is one of the most current studies of a teacher incentive program implemented in the United States, and specifically focuses on the Tennessee Value-Added Assessment System (TVAAS). The result of the study showed performance-based pay is not effective when it is connected so strongly to standardized test scores. The case study of TVAAS offers informative results for this paper, because Tennessee is the ground-zero from which VAA began, and VAA is the calculation tool used to evaluate teachers in the TAP system.

Another article summarizing the faults of the Tennessee Value-Added Assessment System, appeared in *Phi Delta Kappan* in 2004. In this article Gerald Bracey reported the research findings of Haggai Kupermintz of the University of Haifa. One of

---

<sup>130</sup> Olson, 2.

<sup>131</sup> Matthew G. Springer, Dale Ballou, Laura Hamilton, Vi-Nhuan Le, J.R. Lockwood, Daniel F. McCaffrey, Matthew Pepper, and Brian M. Stecher, “Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching,” National Center on Performance Incentives, Project on Incentives in Teaching (September 21, 2010): <http://edlab.tc.columbia.edu/files/pointstudy.pdf> (accessed October 22, 2010), 36.

the major faults found by Kupermintz is the way the effectiveness of a teacher is defined by differences in student learning. He explained that classifying the effectiveness of a teacher this way assumes that a teacher's effects are "independent, additive, and linear," almost as if the teacher is in their own "ecosphere," ignoring any value in "collaborations, team teaching, interdisciplinary curricula, student autonomy and active participation in educational decisions..."<sup>132</sup> The Value-Added Assessment used in the TAP system attempts to deal with this issue by looking at classroom achievement gains of students and school-wide achievement gains of students. The problem is that the achievement gains are only measured according to the California Standards Tests in English and math, and therefore ignore other forms of collaboration, team teaching, interdisciplinary curricula, student autonomy and active participation in educational decisions. The TAP system evaluates the teacher using a nineteen point rubric which includes many important teaching qualities like collaboration and team teaching, but no student achievement gains directly connected to those qualities are measured.<sup>133</sup>

Another major fault of the Tennessee Value-Added Assessment System that Kupermintz found unwarranted is a procedure that William Sanders uses that results in criterion-referenced interpretations of teacher effects or comparisons of teacher scores across a district, which means that "a weak teacher in a weak system would receive a more favorable rating than that same teacher in a strong system."<sup>134</sup> Kupermintz also disputed a claim by TVAAS developers that their value-added system calculates and adjusts for outside influences on student learning. Ironically, a report written by William

---

<sup>132</sup> Gerald W. Bracey, "Serious Questions About the Tennessee Value-Added Assessment System," *Phi Delta Kappan*, Vol. 85, no. 5 (January 2004), 716.

<sup>133</sup> The TAP system will be explained in great detail with full citations in the next section.

<sup>134</sup> Bracey, 716.

Sanders found that teachers assigned to white students, rather than black students, were 8% more likely to be judged effective, and nearly 11% less likely to receive low effectiveness ratings. The claim that adjustments can account for outside forces is false, because the strength of outside forces, mostly connected to family and community variables, is too strong and unpredictable to calculate. This was noted in several studies, the most famous being the Coleman Report, also known as the 1966 report titled, “Equality of Educational Opportunity.” Even TVAAS reports show evidence that contradicts their own claims.<sup>135</sup> Without looking carefully into the research, members of Congress, people like Diane Ravitch, school district superintendents, school board members, and teachers are likely to be fooled by the claims of TVAAS, and buy into the adoption of the TAP system which carries with it the Value-Added Assessment system that claims fairness in the way it evaluates teachers.

Another reason why VAA is controversial deals with the ownership and operation of the private companies that are contracted by public schools to help facilitate the VAA system. William Sanders operates the Education Value-Added Assessment System (EVAAS), which is owned by the SAS Institute Software Company, and is contracted by schools to calculate a complex statistical formula for Value-Added Assessment. Sanders created the original value-added system, the Tennessee Value-Added Assessment System (TVAAS), which relies on a formula for calculating Value-Added Assessment. It is so

---

<sup>135</sup> Bracey, 717. Bracey explained that TVAAS data refute TVAAS claims. He wrote, “Kupermintz quotes a 1998 [TVAAS] report that ‘the cumulative gains for schools across the entire state have been found to be unrelated to the racial composition of schools, the percentage of students receiving free and reduced-price lunches, or the mean achievement level of the school.’ This article cites no data source but appears to refer to an unpublished 1997 report whose authors are not identified. The report contains no formal analyses to support the contention, leaving readers only the choice of eyeballing 1,000 scatter plots. Kupermintz was up to that challenge, and he found that schools with more than 90% minority enrollment showed smaller gains in all subject areas. The relationship with poverty was even stronger” 717.

difficult to understand that it would present a challenge for school districts to adopt their own system of calculation, or for other companies to copy the formula. This difficulty almost ensures a monopolistic control over contracts from public schools attempting to use Value-Added Assessment. The SAS Institute Software Company has partnered with TAP, so that whenever TAP is contracted by districts, then the services of SAS are also contracted. TAP was created by Lowell Milken of the Milken Family Foundation, the brother of Michael Milken, the famed billion dollar junk-bond broker. Policymakers should be critically aware of this currently created phenomena where a few individuals have designed a system which requires public education to privately contract with their companies to hold teachers accountable, or at least create a perception that the performance of teachers is being carefully measured.

Much of the research and push for the Value-Added Assessment comes from William Sanders, who offers his findings as a researcher.<sup>136</sup> The supporters of VAA are a small group of names that appear repeatedly in the various sources that deal with the subject. For example, the website of the Education Consumers Clearing House features an article that celebrates value-added assessment calling it a revolution. The website describes itself as a source for “Education Facts and networking for parents, policymakers, and taxpayers.” The article is written by John E. Stone, and it appears at

---

<sup>136</sup> Here is what the SAS Institute Inc. website says about Sanders: Dr. William L. Sanders is a senior research fellow with the University of North Carolina system and is senior manager of value-added assessment and research for SAS Institute Inc. in Cary, N.C. He assumed the SAS position in June of 2000, upon retiring after more than 34 years as professor and director of the University of Tennessee's Value-Added Research and Assessment Center. In July 2006, Sanders testified before the House Education and Workforce Committee about the reauthorization of NCLB. In February 2007, he shared his research in a Senate Health, Education, Labor and Pensions Committee Round Table discussion on teacher incentives. In 2006 the Education Research Center reported the most influential research on the last decade's national educational policy, and Sanders' value added research developed for Tennessee ranked sixth. [http://www.sas.com/govedu/edu/bio\\_sanders.html](http://www.sas.com/govedu/edu/bio_sanders.html) (accessed April 20, 2011).

first glance as an unbiased reporting of research findings. Although, without too much difficulty, a reader can figure out that the article is a large advertisement for EVAAS.<sup>137</sup>

For example, the language found in the conclusion looks as if it would appear in a brochure selling Value-Added Assessment:

Although it employs some complex statistics, value-added assessment creates a simple but enormously important change in the educational landscape. It enables parents, taxpayers, and education decision-makers to see for themselves whether schools are working. It does so by greatly simplifying the process of interpreting reports on school effectiveness. Such a change can revolutionize education. The public has been flooded with information about school quality but making sense of it has required experts and most of the experts have been educators who work for or with the schools. Now schools can produce a balance sheet and report an objective bottom line that is understandable to the interested citizen. Eventually, resources and students will flow to the effective schools and away from the ineffective ones.<sup>138</sup>

Seven of the eight sources listed as references were written by William Sanders along with various co-authors. The source for empirical findings is the University of Tennessee Value-Added Research and Assessment Center, which William Sanders directed for 34 years, and it is where he initially developed the formula for VAA before he took the private corporate position of running the EVAAS division of SAS.

Sycamore River Unified School District is a perfect example of how educational money flows like a triangulated money-funnel. The district received the 7.2 million dollar Teacher Incentive Fund (TIF) grant from the Federal Government. The grant stipulates that TAP be used as the system for which teachers would be held accountable for their incentive pay. Then, the district pays TAP to train their mentor and master

---

<sup>137</sup> John E. Stone, "Value-Added Assessment: An Accountability Revolution" [http://www.education-consumers.com/articles/value\\_added\\_assessment.shtml](http://www.education-consumers.com/articles/value_added_assessment.shtml) (accessed April 20, 2011).

<sup>138</sup> John E. Stone, "Value-Added Assessment: An Accountability Revolution" [http://www.education-consumers.com/articles/value\\_added\\_assessment.shtml](http://www.education-consumers.com/articles/value_added_assessment.shtml) (accessed April 20, 2011).

teachers that will facilitate the evaluation process. The district is also obligated to use SAS to calculate the complicated value-added formula. Sometimes, William Sanders is successful in lobbying a state legislature to make this obligation actual state law.<sup>139</sup> The district merely acts as a middleman for much of its grant funds which go from the Federal Government to these private companies.

The debate over pros and cons of standardized testing, teacher merit pay, and project-based learning seems to cycle over and over through the decades. The Teacher Incentive Fund grant coming from the Federal Government is the latest push for teacher merit pay, but the idea of giving teachers incentives to encourage better teaching is not new. In 1999 Wellford Wilms and Richard Chapleau wrote an article for *Education Week* summarizing how teacher performance pay systems fail.<sup>140</sup> Just like the findings of the 1990 research group, these two researchers found a historical pattern of the call for educational reform being answered with teacher merit pay. Unfortunately, the inevitable failure of incentive programs due to cheating and curriculum narrowing hampered the reform efforts.<sup>141</sup>

These researchers reached back three centuries to 1710, when England implemented the first teacher pay-for-performance system. Later, it was adopted into

---

<sup>139</sup> An example of this state legislation to require use of EVAAS in schools is HB1669 from North Carolina (May 13, 2010). It states, "An Act to require school improvement teams to use EVAAS or compatible system to collect diagnostic information on students and to use that information to improve student achievement as recommended by the joint legislative education oversight committee." This is an example of how people, like William Sanders, have been able to influence Congress and some state legislatures with his Value-Added Assessment system.

<sup>140</sup> Interestingly, Richard Chapleau received the 1995 Milken Educator award, which is a bit of an irony, because he wrote against an idea being pushed by TAP, which is a Milken Family Foundation creation, and is coupled with the Teacher Incentive Fund grant.

<sup>141</sup> From all of the sources included in my research, a repeating pattern regarding teacher merit pay systems showed that these systems do not last long term. For the interest of brevity, this paper cannot provide all of the details of all of these historical systems that have been observed by researchers. The point of this paper, regarding teacher merit pay systems, is to record the details of the theoretical debate over teacher merit pay that has occurred in recent history, specifically Value-Added Assessment systems.

England's long term education policy of 1862, known as the *Revised Education Code*, and remained there for more than 30 years. The teacher incentive system paid teachers based on their students' test scores in reading, writing, and arithmetic. The curriculum was narrowed to focus exclusively on those subjects. Teachers obsessed to the point that the system was dubbed the "cult of the cash register." One teacher from that time explained the teachers' predicament, "I do not deny that many teachers do overwork the youngsters in a terrible way, but the poor souls really act under the pressure of the law of self-preservation. They must either meet the requirement of their superiors or become professionally extinct." One inspector wrote that the Education Code "did all the thinking for the teacher; it told him in precise detail what he was to do each year." Another inspector wrote, "Every teacher in the country takes his orders from the Code, studies the Code, and devotes his energies to satisfy or to circumvent it."<sup>142</sup>

Written 150 years ago, the observations about England's teacher merit pay system could easily be made about teachers in the United States today that are desperately trying to raise test scores so that their "Program Improvement" school is not taken over by a Local Educational Agency (LEA) as defined by NCLB requirements.<sup>143</sup> As the stakes get higher, so too does the temptation to focus exclusively on ways to raise test scores. The pay-for-performance experiment in England's education system ended in the 1890s.

---

<sup>142</sup> Wellford W. Wilms and Richard R. Chapleau, "The Illusion of Paying Teachers for Student Performance," in *Taking Sides: Clashing Views on Controversial Issues in Secondary Education*, ed. Dennis L. Evans (Guilford: McGraw-Hill Companies, Inc., 2002), 192.

<sup>143</sup> According to NCLB Program Improvement School Requirements, "if a school has not met its Annual Yearly Progress (AYP) then a Local Educational Agency (LEA) identifies the school for corrective action and does at least one of the following: replaces school staff, implements new curriculum, decreases management authority at school level, appoints outside expert, extends school year or day, restructures internal organizational structure of school, LEA informs parents and public of corrective action and allows comment. LEAs may provide direct technical assistance to school site councils in developing school plans." <http://www.cde.ca.gov/ta/ac/ti/nclbpireq.asp> (accessed April 27, 2011).

The incentive system was so abused that schools were known as “grant factories,” and students were dubbed “grant-earning units.”<sup>144</sup> Unfortunately, less than a century later, the United States adopted that idea from England that education could be fixed if only teachers worked harder. In 1969, the Nixon administration started the first experiment with performance contracting, and ever since various districts around the country have experimented with it. The Tennessee Value-Added Assessment System and TAP are both examples of these types of experiments. Wilms and Chapleau contended that teacher merit pay has never led to lasting educational reform, and that reform can only come from local pressure. They concluded their article with this statement, “But only teachers, parents, and students working together at the schoolhouse level can improve the systems by which teachers teach and students learn.”<sup>145</sup> This viewpoint stands against standardized testing and encourages local decisions, which allow more emphasis on portfolios and 21st century learning skills. Standards can still be established at the federal and state levels, but student achievement should be assessed locally and can be regulated by a state or regional group in the same way that accreditation audits are conducted.

One of the major motivations behind current educational reform is to make sure that no child is left behind. The issues and concerns detailing debate over educational reform are connected to some very real problems. Various case studies teach critical lessons about what works and what doesn’t work. One valid concern of reformers considers the disparity between lower-performing schools dominated by poverty and higher performing schools from middle class neighborhoods. A study of the North

---

<sup>144</sup> Wilms, 192.

<sup>145</sup> Wilms, 194.



Carolina schools accountability system found that this system caused a lack of willingness of high-quality teachers to stay at lower-performing schools and presented a challenge for those schools to replace the vacancies left by those teachers. An interesting result of the quality teacher retention problem at the lower-performing schools was a willingness of policymakers to stray from the incentive system set up for teachers by simply raising the base-level salary of teachers at the lower-performing schools to encourage quality teachers to stay at those schools.<sup>146</sup> The Sycamore River TAP system attempts to alleviate this problem by only implementing the system in schools with a 50%+ of students with free or reduced lunch, so the bonus opportunity that TAP offers is only open to teachers at these schools.

A case study of educational reform in New Zealand, in the late 1980s and early 1990s, exposed the dangers of competition and how the result can be an overrepresentation of disadvantaged students in unsuccessful schools. The study analyzed the effects of competition and parental choice regarding school-to-school comparisons of student achievement results and found, “The benefits to the schools serving advantaged students are intensified and the problems of the schools serving disadvantaged students are exacerbated.”<sup>147</sup> When competition leads to a separation between advantaged and disadvantaged students, then a system that rewards teachers for student achievement is not equitable. The compounding effect of outside forces over learning will always give an edge to the teacher with more affluent students. This can

---

<sup>146</sup> Charles T. Clotfelter, Helen F. Ladd, Jacob L. Vigdor, and Roger Aliaga Diaz, “Do School Accountability Systems Make It More Difficult for Low-Performing Schools to Attract and Retain High Quality Teachers,” *Journal of Policy Analysis and Management* 23, no. 2 (Spring 2004): <http://www.jstor.org/stable/3326150> (accessed October 15, 2010).

<sup>147</sup> Edward B. Fisk and Helen F. Ladd, *When Schools Compete: A Cautionary Tale* (Washington D.C.: Brookings Institution Press, 2000), 250.

lead to low morale for teachers of disadvantaged students and destructively affect teacher retention and recruitment in some schools.

This lesson is illustrated in a case study published in 2005, that researched the effects of the implications of the accountability testing in place in England for a recent period of twenty-five years. Valuable lessons emerge about the destructive forces of England's national accountability measures. Recruitment and retention of its teachers is at a recently historic low. The study used Finland to show a contrasting system to England's national accountability movement. Instead of using a national accountability system that diminishes local autonomy and control, the study credited Finland's successful education system to the way it encourages the autonomy and respect that teachers are given.<sup>148</sup>

There was a time in the United States when a similar goal of autonomy and respect drove reformers to consider various systems of teacher merit pay that could foster that type of environment. Until recently, there was no concern over the evaluations of teachers for merit pay purposes possessing any connection at all to standardized testing results. The practice of using standardized test scores to evaluate teachers began in the 1990s for the United States. Unfortunately, the ease of administering and calculating results from standardized testing is too tempting an evaluation tool to resist. A collection of essays published in 1985, in *Merit, Money and Teachers' Careers: Studies on Merit Pay and Career Ladders for Teachers*, brought together the findings of researchers who explored various issues around teacher merit pay, none of which connected standardized test results to teacher evaluations. The possible ways to evaluate teachers for merit pay

---

<sup>148</sup> Caroline V. Gipps, "Accountability Testing and the Implications for Teacher Professionalism," in *Measurement and Research in the Accountability Era*, ed. Carol Anne Dwyer (Mahwah: Lawrence Erlbaum Associates, Publishers, 2005) 99-105.

purposes, according to these experts, involves a time consuming, intricate, system of evaluation.<sup>149</sup> This is a different way to approach the philosophy behind teacher merit pay. Instead of creating a system that is based on standardized test results that are largely outside the control of the teacher, the system aims to create an environment that is professionally supportive of the teacher. Their findings include some key considerations of ways to support teachers to encourage effective teaching:

- (1) intrinsic satisfaction derived from contributions made to student achievement;
- (2) self-esteem based on awareness of one's expertise;
- (3) recognition by peers and "relevant others" of professional competence;
- (4) some opportunities for self-direction;
- (5) positive social interactions with peers and supervisors;
- (6) protection from arbitrary use of authority that might threaten job security or possibilities for advancement;
- (7) opportunities for professional growth and development; and
- (8) economic benefits.<sup>150</sup>

In a supportive environment, where high standards define the expectation and the collegiality of the teachers and administrators, the staff works as a team to raise student achievement. Raising student achievement is the end goal to be reached with the support of the team and with other evaluation measures outside of standardized test results. Several of the researchers warned that merit pay systems can work against a team effort to raise a high standard, because "the majority is defined as sub-standard and paid less than the few who attain excellence."<sup>151</sup> The TAP system does incorporate this supportive team environment through its use of a teacher collaboration and coaching system, and

---

<sup>149</sup> Henry C. Johnson, ed., *Merit, Money and Teachers' Careers: Studies on Merit Pay and Career Ladders for Teachers*. Lanham: University Press of America, Inc., 1985.

<sup>150</sup> William Hawley, "The Limits and Potential of Performance-Based Pay as a Source of School Improvement," in *Merit, Money and Teachers' Careers: Studies on Merit Pay and Career Ladders for Teachers*, ed. Henry C. Johnson, Jr. (Lanham: University Press of America, Inc., 1985) 12.

<sup>151</sup> Terry Herndon, "Merit Pay and the Concerns of the Teaching Profession," In *Merit, Money and Teachers' Careers: Studies on Merit Pay and Career Ladders for Teachers*, ed. Henry C. Johnson, Jr., (Lanham: University Press of America, Inc., 1985), 94.

David F. Wood and Dan S. Green, "Merit Pay and the Concerns of the Teaching Profession," In *Merit, Money and Teachers' Careers: Studies on Merit Pay and Career Ladders for Teachers*, ed. Henry C. Johnson, Jr., (Lanham: University Press of America, Inc., 1985), 126-127.

half the teacher evaluation is determined by a rubric made up of quality teaching techniques. However, almost all TAP systems only measure student achievement growth using standardized test results, so if the goal of the system is to increase student achievement, then the whole system breaks down because of the inadequacy of those test results.<sup>152</sup>

Terry Herndon is another researcher who also published her concerns in 1985. She raised concerns regarding merit pay like the abusive game of playing politics that can occur so that favoritism is rewarded over better performance. This can effect TAP evaluations because half of the consideration is outside of standardized test results. She raised another concern that merit pay, as a documented public record, can be used by parents to select which teachers they want and don't want to teach their kids. This should be an equity concern for TAP systems, because standardized test results are determined to a large extent by outside measures that the teacher has no control over. If a teacher unfairly misses the criteria necessary to receive merit pay, then that teacher is at risk of discrimination by community members. Another concern Herndon raised is the way that merit pay can stifle extra effort by teachers who focus their efforts exclusively on the criteria needed to meet merit pay requirements, because as she stated, outside of a merit pay system excellent teachers will go above and beyond their job description for a variety of personal, social, or spiritual reasons.<sup>153</sup>

From that collection of essays published in 1985, in *Merit, Money and Teachers' Careers: Studies on Merit Pay and Career Ladders for Teachers*, two researchers reported that the teacher merit pay system that they analyzed was too complicated and

---

<sup>152</sup> The TAP system will be explained in greater detail in the next section.

<sup>153</sup> Herndon, 94-96.

burdensome to work effectively long-term. Linda Dockery and Marcia Epstein studied the Teacher Incentive Project (TIP) that was implemented in Winston/Salem, December, 1983-August, 1984. It was a noble attempt to make teacher merit pay work without using standardized test scores as the determining criteria for evaluating teacher performance. Instead, the project involved a teacher video taping himself or herself for twelve hours, and then selecting the best two hours to submit to a board of evaluators, some administrators and some teacher evaluators. The implementation of the project could be broken down into twelve steps: 1) Identification of Principals to serve as evaluators; 2) Identification of Teachers to participate in the piloted project; 3) Identification of Consortium Teams; 4) Identification of Consulting Teachers (they provide training and technical assistance to principals to improve the evaluation function); 5) Project Orientation to train all project personnel--principals, evaluatees, consortium teams, and consulting teachers; 6) Videotaping of classroom teaching; 7) Rating of teachers; 8) Collection and analysis of data; 9) Technical assistance in cases of significant divergence in evaluation results; 10) Post training teacher evaluation with re-evaluations if necessary; 11) Selection of teachers for incentives.<sup>154</sup> This time-consuming, complicated, intricate process simply could not be effectively implemented. This study from the early 1980s demonstrates why educational reform faces a problem today. At a time when standardized testing should be de-emphasized, teacher merit pay has come back into vogue as the latest educational reform. Educational policymakers today fall into the temptation to use simpler means to evaluate teacher performance for merit pay,

---

<sup>154</sup> Linda Dockery and Marcia Epstein, "The Teacher Incentive Project (TIP) of the Winston-Salem/Forsyth County Schools," In *Merit, Money and Teachers' Careers: Studies on Merit Pay and Career Ladders for Teachers*, ed. Henry C. Johnson, Jr., (Lanham: University Press of America, Inc., 1985), 212.

which means that many districts choose to use standardized test scores, breathing new life into the whole standardized testing movement.

To warn against this temptation, the Economic Policy Institute published a briefing paper in August 2010, co-authored by ten leading experts in the field of educational policy research.<sup>155</sup> Together, the authors combined the findings of multiple studies dealing with value-added models and the use of standardized testing results to determine teacher merit pay. Many of these experts lead professional academies, councils, and associations directly dealing with issues involving educational assessment, like the National Center for Evaluation Standards and Student Teaching (CRESST), the National Council on Measurement in Education, the National Research Council's Board on Testing and Assessment, and the American Educational Research Association. Together they represent volumes of research findings, which earns them credibility to collectively

---

<sup>155</sup> These ten researchers are among the most respected experts in the educational field. The reason this long description of their credentials is footnoted in this paper is to emphasize the magnitude of their expertise. The following information quotes the briefing paper's description of the expertise of each co-author: **Eva L. Baker** is professor of education at UCLA, co-director of the National Center for Evaluation Standards and Student Testing (CRESST), and co-chaired the committee to revise testing standards of the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education. **Paul E. Barton** is the former director of the Policy Information Center of the Educational Testing Service and associate director of the National Assessment of Educational Progress. **Linda Darling-Hammond** is a professor of education at Stanford University, former president of the American Educational Research Association, and a member of the National Academy of Education. **Edward Haertel** is a professor of education at Stanford University, former president of the National Council on Measurement in Education, Chair of the National Research Council's Board on Testing and Assessment, and a former chair of the committee on methodology of the National Assessment Governing Board. **Helen F. Ladd** is professor of Public Policy and Economics at Duke University and president-elect of the Association for Public Policy Analysis and Management. **Robert L. Linn** is a distinguished professor emeritus at the University of Colorado, and has served as president of the National Council on Measurement in Education and of the American Educational Research Association, and as chair of the National Research Council's Board on Testing and Assessment. **Diane Ravitch** is a research professor at New York University and historian of American education. **Richard Rothstein** is a research associate of the Economic Policy Institute. **Richard J. Shavelson** is a professor of education (emeritus) at Stanford University and former president of the American Educational Research Association. **Lorrie A. Shepard** is dean and professor, School of Education, University of Colorado at Boulder, a former president of the American Educational Research Association, and the immediate past president of the National Academy of Education.

warn educational policymakers of the potential consequences of using standardized test results to make retention and salary decisions about teachers. They broke down all the reasons why this is the case, from summer retention loss, to less teacher collaboration, to narrowing the curriculum. All the previously discussed research findings in this paper that argue against using standardized tests to evaluate teachers are confirmed by these experts.

The findings of these experts warned policymakers that teacher merit pay systems that heavily rely on standardized testing results are not supported by research. They cautioned that the evidence proves that teacher merit pay systems (like TAP) “which give as much as 50% of the weight in teacher evaluation and compensation decisions to scores on existing tests of basic skills in math and reading...is unwise.”<sup>156</sup> While they admitted that value-added models are a better evaluation of teachers than a single snapshot test score, they report, “The research community has cautioned against the heavy reliance on test scores, even when sophisticated VAM methods are used, for high-stakes decisions such as pay, evaluation, or tenure.”<sup>157</sup> They provided an example of an official statement by the Board on Testing and Assessment of the National Research Council of the National Academy of Sciences that stated, “...VAM estimates of teacher effectiveness should not be used to make operational decisions because such estimates are far too unstable to be considered fair or reliable.”<sup>158</sup>

---

<sup>156</sup> Eva L. Baker, Paul E. Barton, Linda Darling-Hammond, Edward Haertel, Helen F. Ladd, Robert L. Linn, Diane Ravitch, Richard Rothstein, Richard J. Shavelson, and Lorrie A. Shepard, “Problems with the Use of Student Test Scores to Evaluate Teachers,” *Economic Policy Institute, Briefing Paper #278* (August 29, 2010): [http://epi.3cdn.net/724cd9a1eb91c40ff0\\_hwm6iij90.p](http://epi.3cdn.net/724cd9a1eb91c40ff0_hwm6iij90.p) (accessed October 20, 2010), 2.

<sup>157</sup> Baker, “Problems with the Use of Student Test Scores to Evaluate Teachers,” 7-9.

<sup>158</sup> Baker, “Problems with the Use of Student Test Scores to Evaluate Teachers,” 2.

The reasons the VAM results are not considered fair or reliable is because the findings of researchers show that there are many influences over learning that are outside the control of the teacher like: previous teachers, school attendance, out-of-school learning experiences, parental support, family resources, student health, family mobility, the influence of neighborhood peers and/or classmates, and summer learning loss.<sup>159</sup> Influences outside the teacher's control works in both directions. High levels of educational and socioeconomic backgrounds of parents and communities promote learning outside of the classroom with extra reading and life-experiences like visiting a museum. Whereas, students from parents and communities defined by poverty and lack of education not only receive less support to learn outside the classroom, but their learning in the classroom can be severely affected depending on their diet, sleep habits, emotional well-being, etc. They report, "Three-fourths of schools identified as being in the bottom 20% of all schools, based on the scores of students during the school year, would not be so identified if differences in learning outside of school were taken into account." They also show that teachers who teach English Language Learners, special education students, and low-income students are found to receive lower "effectiveness" scores.<sup>160</sup> According to these experts and the research findings that they summarized, the claim that value-added methods can calculate and adjust for student demographic characteristics is overstated and the measures used by VAM are too unstable to be used for the evaluation of instruction or teachers.<sup>161</sup>

A Policy Information Report published by the Educational Testing Service gave a similar warning that there are too many outside influences affecting student performance

---

<sup>159</sup> Baker, "Problems with the Use of Student Test Scores to Evaluate Teachers," 8-17.

<sup>160</sup> Baker, "Problems with the Use of Student Test Scores to Evaluate Teachers," 3.

<sup>161</sup> Baker, "Problems with the Use of Student Test Scores to Evaluate Teachers," 7-9.



to create high-stakes teacher evaluations based on student performance on standardized test results. The report explained how school violence and discipline (things that are often outside the range of control by that one teacher) can adversely affect the test scores. The report stated, "...the issue of school disorder is more than a security and safety problem--it is a critical factor in student achievement. Without order in our classrooms, teachers can't teach and students can't learn."<sup>162</sup> School discipline policy is ultimately made by administrators, not teachers. Why should a teacher's performance be evaluated according to a snapshot-multiple choice test affected by policies created and implemented outside the classroom?

Does test-based accountability increase learning gains? The experts that co-authored the Economic Policy Institute Briefing Paper looked to NCLB results to answer this question, and they found that in many ways NCLB has slowed annual student achievement gains. Before NCLB introduced negative consequences for low test scores, the National Assessment of Educational Progress (NAEP) tested a random selection of students to collect data, but not to penalize any schools. The National Center for Education Statistics explains how NAEP is a *Common Yardstick*:

The National Assessment of Educational Progress (NAEP) is the largest nationally representative and continuing assessment of what America's students know and can do in various subject areas. Assessments are conducted periodically in mathematics, reading, science, writing, the arts, civics, economics, geography, and U.S. history.

Since NAEP assessments are administered uniformly using the same sets of test booklets across the nation, NAEP results serve as a common metric for all states and selected urban districts. The assessment stays essentially the same from year

---

<sup>162</sup> Paul E. Barton, *A Policy Information Report--Order in the Classroom: Violence, Discipline, and Student Achievement* (Princeton: Educational Testing Service, 1998), 46.

to year, with only carefully documented changes. This permits NAEP to provide a clear picture of student academic progress over time.<sup>163</sup>

It is important to note that the testing under NAEP carried no accountability with it, so instruction was not influenced by any positive or negative incentives for teachers or schools. Whereas, NCLB was designed to hold teachers and schools accountable and carried with it both positive and negative incentives. From comparing NAEP test results (1992-2003) to NCLB test results (2003-2009), and focusing on average annual rates of test-score growth for African Americans and white students, the experts showed that high-stakes testing accountability, implemented by NCLB, has not produced the positive educational reform that so many NCLB supporters claimed it would. It has negatively affected education. The data shows that NCLB has slowed the rate of growth in six of the eight areas:

African American students .....	Fourth grade math-	went from 2.2 to 1.0	(-)
African American students .....	Fourth grade reading-	went from 0.5 to 1.1	(+)
African American students .....	Eighth grade math-	went from 1.2 to 1.4	(+)
African American students .....	Fourth grade reading-	went from 0.6 to 0.3	(-)
White students.....	Fourth grade math-	went from 1.8 to 0.8	(-)
White students .....	Fourth grade reading-	went from 0.4 to 0.3	(-)
White students.....	Eighth grade math-	went from 1.4 to 0.9	(-)
White students.....	Eighth grade reading-	went from 0.5 to 0.1	(-) <sup>164</sup>

The warnings from this Economic Policy Briefing Paper follow a progression of recent educational reform efforts that emphasize accountability through standardized testing and teacher merit pay connected to those test results. Both Republican and Democratic presidential administrations under Bush and Obama have embraced the idea that testing accountability measures will raise gains in student achievement. In the face of this baseless enthusiasm from policymakers, the educational experts warn that there

<sup>163</sup> The National Center for Education Statistics <http://nces.ed.gov/nationsreportcard/about/> (accessed June 24, 2011).

<sup>164</sup> Baker, "Problems with the Use of Student Test Scores to Evaluate Teachers," 6.

are reasons to be “skeptical of claims that measuring teachers’ effectiveness by student test scores will lead to the desired outcomes.”<sup>165</sup> One of those reasons is the negative test results from NCLB, stated above. Another reason is the failing track record that incentive pay has demonstrated in both the public and private human service sectors in both the United States and Great Britain. The same sort of cheating that has been observed among teachers and administrators with accountability testing has been observed among public and private sector personnel. Cardiac surgeons, for example, when patient-survival-rates were recorded and published by government sources, were found to turn away the sickest patients. Another reason the educational experts are skeptical of standardized testing and teacher merit pay is the way money for merit pay can be pulled from general funding for teachers, which can result in the stagnation of teacher salaries leading to a negative affect on teacher recruitment and retention efforts. Teachers who are locked out of the merit pay, for various reasons, will experience a delay of a pay raise in the general salary schedule, because the money needed to give a general raise is instead diverted to teacher performance bonuses. Supporters of TAP claim the money will not come from the general budget, because it comes from federal grant and Title I and Title II funding, but money is money, and when the Federal Government is awarding TIF grants, it has less money in its general education budget, which eventually does affect general budgets at the state and local levels. This makes the teaching profession even less competitive for recruitment of qualified professionals. This will become worse in tight fiscal environments and as prospective teachers and existing teachers become more aware of inaccuracy and unfairness of value-added measurements.

---

<sup>165</sup> Baker, “Problems with the Use of Student Test Scores to Evaluate Teachers,” 7.

The final reason for skepticism stated by the experts is largely in line with the focus of this paper. They stated that standardized tests now in place are “narrow measures of what students know and can do, relying largely on multiple choice items that do not evaluate students’ communication skills, depth of knowledge and understanding, or critical thinking and performance abilities.”<sup>166</sup> As the United States pushed to raise scores on these multiple choice tests that access low-level thinking, it slipped in the years from 2000 to 2006 in international ranking on tests that assess more complex skills.<sup>167</sup> The research showed that these standardized tests, mandated by NCLB and used as a way to determine teacher merit pay, narrow curriculum, encourage cheating, and are not “accurate measures of the knowledge that the tests are supposed to measure,” so the result is the “continuing need for remedial courses in universities for high school graduates who scored well on standardized tests, yet still cannot read, write, or calculate well enough for first-year college courses.”<sup>168</sup> The experts did admit that there are many problems with the current system of teacher evaluations, but they conclusively found that the use of teacher merit pay systems, that heavily rely on standardized test results, do not work.

The results of decades of research are starting to become more widely disseminated within the field of education and even into the general ranks of society, placing the research findings more directly in front of educators, parents, and voters, so they access it without digging for it. The push-back against policymakers is finally beginning. Teachers’ unions are feeling the threat enough to devote large portions of their publications to these problems. The *neatoday* magazine issues published during the

---

<sup>166</sup> Baker, “Problems with the Use of Student Test Scores to Evaluate Teachers,” 7.

<sup>167</sup> Baker, “Problems with the Use of Student Test Scores to Evaluate Teachers,” 7.

<sup>168</sup> Baker, “Problems with the Use of Student Test Scores to Evaluate Teachers,” 7.

2010/2011 school year all featured articles dealing with standardized testing results and merit pay. *California Educator* has also attempted to make teachers more aware of these issues. Both these magazine are distributed to nearly every educator in California, and *neatoday* to nearly every educator in the nation. The movie *Race to Nowhere*, released in 2010, received a full page write-up detailing the movie's message and, as of February 2011, 1400 private screenings had occurred in 48 states and 15 countries. The movie "focuses on the mental and physical toll today's competitive school system takes on students, teachers and families due to an overemphasis on testing, drill-and-kill instruction and overwhelming amounts of homework."<sup>169</sup> A November 2009 issue of *Educational Leadership*, a magazine with a target audience of educational administrators, published a total of nine articles dealing with the issue of standardized testing not being an accurate or authentic measure of student achievement. Several of the articles dealt with the issue of connecting standardized test results to teacher merit pay. One author asked a critical question that deserves an answer. She asked, how are policy decisions made when they have not been through a rigorous process of testing like the process that new pharmaceutical drugs have to go through under the FDA? Then, she asked more specifically why is Value-Added Assessment embraced by policymakers as a cure to the ills in education when the research that currently exists raises serious doubts about its effectiveness and serious cautions about its destructiveness?<sup>170</sup> In another article, the author explained how a performance-based system that is strongly tied to standardized test scores is flawed, "A system that rewards schools, students, and teachers only for test

---

<sup>169</sup> *California Educator*: California Teachers Association, 15, no. 5 (February 2011): 18.

<sup>170</sup> Audrey Amrein-Beardsley, "Value-Added Tests: Buyer, Be Aware," *Educational Leadership* 67, no. 3 (November, 2009): 37-42.

scores will get mostly test scores.”<sup>171</sup> The author finished his article by explaining what creates this kind of system: the assumption that teachers lack motivation, the assumption that schools are failing, and the assumption that measuring academic achievement is all that counts. Finally, all of this information is making its way from the researchers into the hands of teachers and others. It is hoped that this concerted awareness will make a united difference.

Unfortunately, policymakers do not get it. Many of them possess the right intentions, but for whatever reason, they push against a torrent of research. For decades, educational research findings have conclusively shown that standardized testing is not accurate, and not authentic, and detracts from learning important skills like critical thinking, collaboration, communication, and self-directed learning. For these reasons, standardized test results should not be used to evaluate teachers. Merit pay in whatever form has a proven track-record of failure.

The debate should be over. Why do policymakers push against educational research? It seems that it is a matter of ideology versus evidence. In this case, the policymakers are driven by ideology more than research findings. At a time when the public is starting to understand the findings of educational research, policymakers ratchet up their push for standardized test results by embracing teacher merit pay systems that rely heavily on those results. At the same time all of this is happening, corporate leaders explain that 21st century learning skills must be taught, and practiced, and learned in order for the United States to remain competitive in the world. It is beyond the scope of this paper, but further research could provide insight into how purposefully this

---

<sup>171</sup> Donald B. Gratz, “The Problem with Performance Pay: It can work, but only if performance is broadly defined and all parties agree to the plan,” *Educational Leadership* 67, no. 3 (November, 2009): 79.

disconnect has been planned and implemented. Some groups driven by conservative ideology want to take down public education. There is a push for privatizing education through the voucher system. There is a push to dissolve teachers' unions. These agendas seem like they are gaining momentum in recent decades. Is that why the debate continues? Could it be that some conservatives push policy that encourage classroom minutes to be spent on irrelevant learning so that more students leave school incompetent, damning public education? If that's correct, why do liberal policymakers vote for the same policies? *No Child Left Behind* was born and continues to live by bipartisan efforts.

Is it simply that the American educational mission has never been settled, and that the debates by Thorndike and Dewey continue to live almost a century later. Diane Ravitch pondered this question in 1977, in a book she wrote about the historiography of American education. She posed the idea that maybe "schools do not have cosmic purposes; that they cannot 'save' society; that they are neither spearheads of radical change nor instruments of cultural repression," and then to reframe the way we think about education she encouraged her readers to, "think instead of institutions whose purposes are circumscribed by the public that supports them, and whose goals are limited and potentially attainable."<sup>172</sup> If society pushes for standardized testing results, then somehow the schooling system will learn a way to raise the scores, but is that really the goal that society wants?

What happened to recommendations like those from the Educational Policies Commission in the 1930s that wanted to encourage students to maximize the power of Democracy through the ability to critically think and communicate:

---

<sup>172</sup> Diane Ravitch, *The Revisionists Revised: A Critique of the Radical Attack on the Schools* (New York: Basic Books, Inc., 1978), 172.

Most of the standardized testing instruments and written examinations used in school today deal largely with information....There should be a much greater concern with the development of attitudes, interests, ideals, and habits. To focus tests exclusively on the acquisition and retention of information may recognize objectives of education which are relatively unimportant. Measuring the results of education must be increasingly concerned with such questions as these: Are the children growing in their ability to work together for a common end? Do they show greater skill in collecting and weighing evidence? Are they learning to be fair and tolerant in situations where conflicts arise? Are they sympathetic in the presence of suffering and indignant in the presence of injustice? Do they show greater concern about questions of civic, social, and economic importance?<sup>173</sup>

These ideals never left many of the classrooms across America, because these ideals never left America.<sup>174</sup> Even in the midst of all the multiple choice tests, society still wants more from schools. Whatever causes the mixed messages, the present policies push and pull teachers in multiple directions leaving them with a certain feeling of schizophrenia. Despite how policymakers change their agendas, many veteran teachers over the years have learned to survive through the pendulum swings of reform. They bobble their head up and down in staff meetings, and then return to their classroom, close the door, and teach in ways that they know work, because they stand on solid ground formed by thousands of hours of experience and supported by research. They know that educational reform of one sort or another comes and goes, first phonics, then whole language, then phonics again. The talented teachers constantly challenge themselves to try new things, adopting what works and dumping the rest, always refining, refining, refining. Tragically, in today's educational climate, it is becoming increasingly difficult to be a talented teacher. Like soldiers marching, policymakers call the cadence and teachers either stay instep or fall out.

---

<sup>173</sup> Rothstein, Jacobsen, and Wilder, 24-25.

<sup>174</sup> Some people driven by their own personal greed have tried to label the idea of *working together for a common end* as a socialist or a communist idea. Actually, you do not have to be a historian to understand that many of the greatest American achievements have been through *people working together for a common end*.



After studying the educational research, I wonder how many teachers understand how far off the policymakers are from the research findings. I wonder how many teachers are still trying to be selective and do the best teaching they know how, or how many are just walking in stride with the mania of *No Child Left Behind* or the latest push for teacher merit pay that relies on standardized tests and Value-Added Assessment. I wonder how many realize what has happened to education in the last twenty years, how we have dropped vocational education and the arts and anything else that doesn't somehow raise the scores on the high-stakes tests. I'm sure it is like a parent who gradually sees their child growing, but doesn't truly appreciate the size difference until a friend points it out.

I once had an experience in the Army on an eighteen mile road-march back to the garrison after two very tiring weeks of field training. The eighty-five soldiers in my unit walked single file along the road with a full rucksack on our back and fifteen feet separating each person. We walked for six hours. Every fifty minutes we would stop and rest for ten minutes. For one of those fifty minute time periods I fell asleep, but I continued to walk. I did not realize it until we stopped again. I thought for sure we had just stopped. The time seemed to pass so fast, because I was not conscious. I asked my instructor about this, and he said it happens, that sometimes soldiers in tip top shape will fall asleep on a long road-march, and yet they will keep in line and keep their pacing perfectly.

I wonder how many teachers are asleep on this march of standardized testing.

## Chapter V

### The TAP System, the Direction of Education in Sycamore River and the Nation, and Relative Teacher Perceptions

The home webpage of TAP claims that this system encourages teacher excellence, student achievement, and opportunities for all, and that the system was designed “to attract, develop, motivate and retain the highly effective educators that all students deserve. In 2005, Lowell Milken established the National Institute for Excellence in Teaching to manage and support TAP's effectiveness, growth and sustainability system.”<sup>175</sup> The superintendent of Sycamore River explained what attracted him to TAP:

My motive for wanting to bring the TAP system to Sycamore River was two-fold. First of all, I wanted to bring an instructional model to our district that is currently lacking. I first read about TAP's 19 point rubric, in a report entitled "Aligned by Design," and discovered it was based on a number of frameworks such as Charlotte Danielson's Framework for Teaching and the University of Virginia's Classroom Assessment Scoring System, among others. While not perfect by any means, the rubric provides a written description of specific teacher competencies that is based on a 5 point scale, with a "3" representing proficient performance. Secondly, I was excited about the model for using master and mentor teachers to help provide on-site feedback and coaching on a daily basis to help our teachers improve. Weekly cluster meetings are led by the master and mentor teachers in a professional learning community format. My most basic intent in applying for the TIF grant to pay for the TAP system was simply to provide the district with a structure and supports that could help our teachers and our principals to refine their professional skills.<sup>176</sup>

TAP is a system that allows for multiple career paths so teachers can step into leadership and increase their pay as a master or a mentor teacher. This system fosters ongoing professional development of about an hour a week for cluster meetings where teachers can collaborate and receive coaching from the mentor and master teachers on the

---

<sup>175</sup> More information about what TAP claims can be found at <http://www.tapsystem.org/> (accessed June 14, 2011).

<sup>176</sup> Superintendent News, October 1, 2010. The full address to the district staff can be viewed at the district website, but as mentioned in the introduction, this information is purposefully deleted to protect individuals and the local school district that is being studied (accessed June 14, 2011).

most effective teaching techniques. The coaching continues through instructionally focused accountability where teachers are observed at least four times a year. TAP teachers are rewarded with performance-based compensation of a possible \$3,000 annually.<sup>177</sup> According to the Sycamore River negotiated TAP agreement with its teachers, “Performance awards are based on three criteria: 1) The final SKR Score in the teacher’s summative evaluation [evaluations based off the nineteen point rubric listed below], 2) Classroom achievement gains, 3) School achievement gains.”<sup>178</sup> The breakdown of these three categories differ slightly depending on what grade level and what subject each teacher teaches. Half the performance for all TAP teachers in Sycamore River will be determined according to Value-Added Assessment based off the California Standards Test (CST). For English and math teachers, part of the percentage of their performance will be based off classroom achievement gains, and for all other teachers it will be based off school achievement gains. *All the achievement gains are measured using the CST.*<sup>179</sup> The quality of instruction is driven by a nineteen point rubric:

1) Standards and Objectives, 2) Motivating Students, 3) Presenting Instructional Content 4) Lesson Structures and Pacing, 5) Activities and Materials, 6) Questioning, 7) Academic Feedback, 8) Grouping Students, 9) Teacher Content Knowledge, 10) Teacher Knowledge of Students, 11) Thinking, 12) Problem Solving, 13) Instructional Plans, 14) Student Work, 15) Assessment, 16) Expectations, 17) Managing Student Behavior, 18) Environment, 19) Respectful Culture.<sup>180</sup>

---

<sup>177</sup> A more detailed description of TAP by the superintendent can be found in Appendix D.

<sup>178</sup> Sycamore River TAP Agreement--details were found on the real district website [deleted] under School Board documents: Sycamore River Unified School District Board of Education Agenda #1, February 1, 2011, Public Hearing, AB 1200 Disclosure of Costs of Sycamore River Unified Teachers Association Tentative Agreement (TA), July 1, 2010, through June 30, 2012. (accessed June 14, 2011).

<sup>179</sup> Sycamore River TAP Agreement

<sup>180</sup> This rubric is apart of the Teacher Advancement Program Leadership Handbook, (193-200). The rubric was originally attached to an email from Sycamore River’s superintendent, October 1, 2010. I cannot find the rubric on any TAP website or the district’s website page.

The details of this rubric combine to create quality instruction. Lowell Milken did not create anything new with these descriptors of effective teaching. All of these ideas have been around for a long time, but by placing them into an organized system of teacher collaboration and coaching, he created a system that can facilitate a school-wide effort of improving instruction in one way or another for all teachers. TAP's steps for effective learning in the classroom are: 1) Identify the problem or need based off a student pre-test, 2) Obtain new teacher learning aligned to student need and formatted for classroom application, 3) Develop new teacher learning with support in the classroom, 4) Apply new teacher learning to the classroom, 5) Evaluate the impact on student performance through a post-test.<sup>181</sup> The direction of instruction for TAP is critical, because where you aim is probably where you will hit. What is the ultimate goal? When the learning outcome is measured by the California Standardized Tests as it is in Sycamore River, then the instruction will be tailored for that outcome. If teachers don't teach with the test as their goal, then they are set up for performance-based failure.

The TAP system is a great system built off sound teaching practices, teacher collaboration, guidance and support, and student data. However, there are several problems I have with TAP. As a national leader in education, why isn't TAP actively speaking out against standardized testing? Surely, the individuals that make up TAP have done the similar surveys of the findings of research experts that went into writing this paper. What reasons do Lowell Milken and his associates have for not devoting energy to change the obsession that policymakers have regarding standardized testing?

---

<sup>181</sup> This information is from a Power Point presentation by a TAP representative given in Sycamore River on November 12, 2010. Information similar to this can be found at <http://www.tapsystem.org/action/action.taf?page=ifa> (accessed June 13, 2011).

Instead, they have designed a system that agrees with standardized testing. Do they force schools that use TAP to measure student achievement using standardized test results?

No, according to the TAP representative that came to Sycamore River, November 12, 2010. That TAP representative explained that most schools choose to use the standardized test results, because they are already measurement tools that are in place and they are easy to use. Sycamore River will use standardized tests to measure student achievement, but only student achievement in English language arts and math. The reason Sycamore River chose such narrow perimeters is probably because of NCLB and the Program Improvement many of its schools face, and because of the ease of this already-in-place assessment tool. As far as my research revealed, nowhere does TAP warn that standardized tests are not accurate or authentic measurements for the type of learning that should be occurring in the classroom.

Why is TAP okay with schools using standardized tests as a measurement of student achievement? Is it because of the consequences schools face regarding NCLB requirements? Why is TAP not visibly crusading to modify NCLB requirements? Why did William Sanders, the creator VAA--the computation tool used by TAP, testify in front of Congress regarding the extension of NCLB? Does TAP remain silent to avoid agitating the same government individuals that make budgeting decisions that result in millions of dollars in grant money that ends up going to TAP? Is it simply a notion of federal power with TAP being fueled by the same power as NCLB--the sacred cow of federal power over education? Strangely, TAP evaluates teachers to see if they are teaching 21st century learning skills, but the TAP system that uses standardized test results does not evaluate student achievement to see if the students have learned the 21st

century learning skills, because those skills generally cannot be measured on standardized tests. The TAP system that is supposed to be driven by data is weakened by missing gaps regarding results in these skills that are so crucial for success today.

TAP possesses limitations, and is only one part of the future direction of Sycamore River along with a push for more project-based learning at all schools. There exists a push to raise scores in Sycamore River to avoid NCLB sanctions, while at the same time, the district is planning real 21st century reforms. One reform is the International Baccalaureate (IB) program that will be implemented at a Sycamore River elementary school in the fall of 2011. This program uses inquiry methodology to teach 21st century skills like self-directed learning and collaboration to work through six trans-disciplinary themes within units of inquiry that last several weeks each. This is how the IB program describes its elementary school program: “Students in the 21st century are faced with the challenge of learning about an interconnected world where knowledge is constantly developing. The International Baccalaureate® (IB) Primary Years Programme prepares students to be active participants in a lifelong journey of learning.”<sup>182</sup>

Another Sycamore River reform effort is the New Tech High that is scheduled to open in the fall of 2012 on the campus of one of Sycamore River’s high schools. This is how New Tech Network describes the beginning of its history, “We began in the mid-90s in Napa, California. The local schools were meeting education standards, and the community thought of Napa High School as a good school. However, local business leaders remained concerned that meeting basic standards would not be enough to ensure that students were graduating with the skills needed to meet the needs of the new

---

<sup>182</sup> The website of the International Baccalaureate Program is <http://www.ibo.org/pyp/> (accessed June 18, 2011).

economy.”<sup>183</sup> Today, New Tech Network is affiliated with 62 public high schools in 14 states. I had the opportunity to visit a New Tech High located near Austin, Texas, along with a small group of Sycamore River teachers, administrators, and school board members, December 13, 2010. Manor New Tech High School uses a TAP system within a STEM (Science, Technology, Engineering, Math) school which focuses exclusively on project-based learning. This is the way Manor New Tech High School describes itself:

The goal of Manor New Technology High School is to prepare students to excel in an information-based and technologically-advanced society. We are committed to leading educational reform and our instructional program encourages students to learn through collaboration with peers, businesses, and the community. Students develop problem-solving skills, interpersonal skills, and the resiliency they need to succeed in a rapidly-changing and competitive world. Our curriculum brings together the strength of modern technology, community partnerships, problem solving, interdisciplinary instruction, and global perspectives in a student-centered, collaborative, project-based community.<sup>184</sup>

The school-wide student to computer ratio is one to one. New Tech’s philosophy is to combine the teacher coaching and collaboration, through the TAP system, with 21st century learning skills practiced in a technology rich, project-based learning environment with projects that are designed around targeted state standards. Manor New Tech High School boasts an impressive record of raised test scores, raised graduation rates, and lowered dropout rates. I believe the success of Manor New Tech comes down to the school-wide culture of project-based learning and the student enthusiasm for such an environment. On average, each student at Manor New Tech presents 65 public speeches annually. The enrollment is 325 students and is decided by a blind lottery, so the diverse group of students who attend share the desire to be in a technologically driven project-

---

<sup>183</sup> The website of New Tech Network is <http://www.newtechnetwork.org/> (accessed June 17, 2011).

<sup>184</sup> The website of Manor New Tech High School is <http://www.manorisd.net/portal/newtech/> (accessed June 17, 2011).

based learning environment. It is unclear how much the enthusiastic student-buy-in to this program boosts standardized test results, because of how seriously the students approach the tests in the name of school pride, or how much the project-based learning environment leads to higher test scores.

The New Tech High approach is expensive. It will cost Sycamore River \$450,000 to pay a start up cost to New Tech Network over the course of 4-5 years, along with the extra cost of one computer for every student in every class, and supporting a small school with an extra administrator and a dedicated technology support person. Sycamore River is approaching the effort with the use of a fund raising effort called the Sycamore River Foundation for Innovation. Even after receiving private donations and grants, Manor New Tech High is now facing the challenge of replacing its first generation of computers. In many ways the investment in Manor New Tech High School is paying off for the small Texas school district, because of the way this special high school is spreading project-based learning to other schools in the district. However, it is not clear how successfully the project-based learning is inculcating into the culture of the other schools. Overall, the New Tech High idea seems very sound, but it is unclear how long or how wide a district can support efforts like this with the present level of funding for education.

New Tech High's expense is due to its heavy reliance on technology and the required support costs from New Tech Network. There is no doubt that technology is important, but if becomes the obstacle that prevents schools from becoming project-based learning schools, then it is best to remember Deborah Meier's experience with the Central Park East schools in Harlem. According to Meier and independent evaluators, the cost of



running a project-based learning school like the primary and secondary schools of Central Park East was comparable to standard public school programs.<sup>185</sup> The student to teacher ratio was twenty to one. The cost for smaller class sizes was offset by cutting administrative overhead (instead, a teacher in charge), or by grouping several similar schools into the same building to share support staff. Teachers teamed with each other and shared the same group of students. Each teacher covered two subjects like math and science, or English and history. The students worked more independently and did not require as much time in a seat in front of the teacher. For Meier, making this type of program financially feasible was how “the children at the bottom of America’s social ladder could use their schools to develop rather than stunt their intellectual potential.”<sup>186</sup> The teachers used *some* direct instruction. The approach attempted to not patronize the students with so much freedom that they ended up missing important knowledge. The teachers taught important ideas rigorously boiled down from the traditional list of state standards. Meier explained, “...we saw children being driven into dumbness by a failure to challenge their curiosity, to build on their natural drive toward competence. We thought adults had important things to teach children, not just a mission to get out of their way.”<sup>187</sup> Financially, there is no reason why other schools cannot use this approach. Why should private school students of wealthier families be the only ones that are able to explore and satisfy their intellectual curiosities?

Meier and the other teachers at Central Park East broke down the walls that hold children in programs that do not work. She used much of her book to show ways to work through the typical excuses that prevent educators and communities from starting project-

---

<sup>185</sup> Deborah Meier, *The Power of Their Ideas: Lessons for America from a Small School in Harlem* (Boston: Beacon Press, 2002), 19.

<sup>186</sup> Meier, 19.

<sup>187</sup> Meier, 21.

based learning programs. For instance, she addressed the issues of racism in school amongst a diversified population. Through active dialogue and collaborative projects, the students learned to respect the talents that each individual possessed. Another obstacle she broke down is parental lack of interest. In Central Park East students were inspired to learn. The parents became curious to see what their children produced. With fewer students, the teachers knew the families of their students through regular timely meetings. The parental support did not happen by accident. The school required it. Meier explained, “We insisted that parents (or grandparents, aunts, older siblings) visit before signing up, and we considered it our job to enlist their collaboration.”<sup>188</sup> Another common excuse for not starting this type of program is that teachers normally cannot agree to work together so closely. Meier explained that it was much easier to settle differences, because the experimentation to find what works was so easy to adjust in a smaller school setting with fewer decision makers that even the students were able to be involved in some of the decision making.<sup>189</sup>

Deborah Meier’s alternative schools, Central Park East, took students from Kindergarten to their Senior year-successfully standing up against standards-based learning and celebrated its first graduation class in 1991. The graduates shattered the normal statistics of New York public school students. Considering that fact that these students were “roughly equivalent to a cross sampling of New York City. The majority of students [were] African American and Latino, most [were] low-income or poor, and they experience[d] a full range of academic strengths and handicaps.”<sup>190</sup> Fewer than five percent who began the program by ninth grade dropped out. The normal dropout rate citywide was half. Even more impressive, ninety percent of the graduates went directly

---

<sup>188</sup> Meier, 22.

<sup>189</sup> Meier, 53-58.

<sup>190</sup> Meier, 16.

to college and stayed there, and the results improved with the program as she wrote her book.<sup>191</sup>

Meier finished her book with a plea to fix public education so this grand experiment of education for all people will not end in failure. For the United States to continue its democracy, it must create an opportunity for students to think, *really* think. Students thrive when they can be creative and engage their mind, and it does not have to be within the confines of an expensive technology rich environment. The students of Central Park East proved it.

My own experience with project-based learning this year proved to me one lesson Meier described, that the success of project-based learning requires a school-wide inculcation and a culture where students grow to understand that the inquiry method is difficult and requires critical thinking and teamwork to work through the challenges that a project presents. If students can learn this important lesson and stick with project-based learning through several projects, then students awake to the excitement of using the creative process to solve problems and the pride of owning and presenting their findings. When project-based learning has not become a part of the school culture, then many students are not comfortable with the self-directed learning and the awkward lack of confidence that creates a stumbling block to really think their way through a scenario without the teacher force-feeding them the answers in lectures or providing worksheets with simple questions. I team-taught several project-based learning classes with another teacher for a semester. We combined two subjects that naturally compliment each other-- American Literature and U.S. History, and facilitated a service project format where students recognized themes of the human condition from the subject matter, and then

---

<sup>191</sup> Meier, 16.

designed and implemented a project with the aim of community improvement. Even though there were many ways we could have improved the class, I still witnessed some of the most exciting learning for students, because they owned their learning, and they experienced the satisfaction of overcoming obstacles to complete a project with real-world importance. I also taught a government and an economics class both formatted exclusively around project-based learning. Despite the challenges, I believe project-based learning needs to be a long lasting reform effort that becomes a part of educational culture so students practice regularly the skills they will need outside of school.

In the midst of all this planned Sycamore River reform for 21st century learning skills and project-based learning, and in the midst of a push to raise test scores, it becomes critical to listen to teachers who day after day, year after year, student after student, conduct their own research in what effective teaching looks like. Sycamore River teachers received an email on May 27, 2011, explaining that I needed Sycamore River teacher input to complete my Master's thesis on the history of standardized testing and teacher merit pay.<sup>192</sup> Within twenty-four hours, I received 100 responses. By June 1, I had received a total of 142 responses.<sup>193</sup> The survey responders represent 37% of the total population of Sycamore River teachers, which is approximately 530 teachers. I also conducted lengthy interviews from May 16-June 4, 2011, of a total of 23 teachers and 2 administrators-all from Sycamore River, except for one high school teacher from another local school district. The teachers from Sycamore River represented nearly all grade levels and nearly all subjects. Each interview lasted anywhere from four to twenty-four minutes. The total time for all the interviews was over four hours. The one high school

---

<sup>192</sup> Appendix A contains the survey results.

<sup>193</sup> A special note of appreciation goes to my brother, Mark Greenelsh, for helping me through the technical process of administering my online survey.

teacher that was outside of the school district expressed the same general concerns as Sycamore River teachers expressed. These interviews were documented on a video camera and recorded onto four DVDs.<sup>194</sup> Many of the comments from these interviews are interjected into this section in block quotes, making up a crucial part of this section that so heavily relies on the voice of local educators.

Survey Question One stated, *NCLB standardized tests do a good job assessing 21st Century Learning Skills*--with four options to choose from (the number in parenthesis shows the amount of responses for each choice): *Strongly Agree (2)*, *Somewhat Agree (27)*, *Somewhat Disagree (44)*, and *Strongly Disagree (69)*. Eighty percent of the teacher respondents answered with Strongly Disagree and Somewhat Disagree. According to this survey, four-fifths of teachers recognize the limitations of standardized tests regarding 21st century learning skills.

While the district moves forward with reform efforts to encourage 21st century learning skills, they place more emphasis on standardized test results by choosing to make the only measurement of student achievement in the TAP system the California Standards Tests. According to Article 14 of the TIF/TAP Negotiations, “Classroom Achievement gains will *only* be based on California Standards tests measured by a third party provider (SAS) for teachers of English/Language Arts and Math in grades 3-8. Classroom Achievement gains shall not be measured for other teachers---including teachers in grades K-2, history, PE, science, special education, and other electives.”<sup>195</sup>

---

<sup>194</sup> Appendix C contains the questions that led the interviews.

<sup>195</sup> TIF/TAP Negotiations Update Sycamore River United Teachers Association emailed document. All of the specific details of the TAP agreement in Sycamore River can be found on the Sycamore River website [actual district website not included for anonymity reasons] under School Board documents: Sycamore River Unified School District Board of Education Agenda #1, February 1, 2011, Public Hearing, AB 1200 Disclosure of Costs of Sycamore River Unified

This means that teachers in Sycamore River are being asked to teach 21st century learning skills through project-based learning, but any student achievement of these skills gained through project-based learning at a TAP site will not be used in evaluating teachers for bonus rewards. It remains to be seen how the notion of *you get what you pay for* will unfold with the TAP system in Sycamore River.

Not only is the focus of student achievement only on standardized tests for performance-based pay, but the subjects within the standardized tests are even further limited. The focus of the measured student achievement strictly limits itself to math and English language arts. One administrator that I interviewed explained that such a narrow focus on math and English language arts is causing education to lose creativity. This administrator felt pressure from NCLB to raise test scores to get out of the Program Improvement status that his or her school was under, but at the same time wanted to use quality project-based learning to teach collaboration, problem solving, thinking-on-your-feet, and being a critical thinker. These types of skills that are learned and practiced within project-based learning are not assessed on standardized tests. This administrator said:

I am an advocate of project-based learning when it is done in the correct way. Well crafted projects that have authentic assessment--I think that's richer learning and bigger picture learning. However, being a principal at a school that is currently in year one Program Improvement, there is pressure to perform, so in order to carry out my job and what is expected of me, I have to shift what I might be passionate about and what I might personally believe is sort of put to the side, and our focus has become to get our scores up so we are not in Program Improvement, so I advocate for that, because I have to, because that's my role right now in this big giant system that has a lot of flaws.<sup>196</sup>

---

Teachers Association Tentative Agreement (TA), July 1, 2010, through June 30, 2012. (accessed June 14, 2011).

<sup>196</sup> The names of the people interviewed will not be reported in this paper in order to protect their identity. Some the teachers interviewed gave their names and other teachers did not. The

According to the TAP instruction rubric, some of the criteria used to evaluate teachers measures how well they teach 21st century learning skills. Teachers are evaluated according to whether they are teaching 21st century learning skills, but no student achievement is measured accordingly. For example, under the section of the rubric titled *Activities and Materials*, it states, “In addition, sometimes activities are game-like, involve simulations, require creating products, and demand self direction and self monitoring.” Under the section of the rubric titled, *Questioning*, teachers are evaluated according to whether their questions lead to “application and analysis, creation and evaluation,” and whether “students generate questions that lead to further inquiry and self-directed learning.”<sup>197</sup> These learning outcomes are considered to be 21st century learning skills, but they are only a small part of the Instruction Rubric used to evaluate teachers under the TAP system.

Survey Question Two stated, *The impact of standardized tests on education in the last 20 years has been--*with four options to choose from: *Very Beneficial (5)*, *Somewhat Beneficial (26)*, *Somewhat Detrimental (67)*, *Very Detrimental (44)*. Seventy-eight percent of Sycamore River teachers surveyed responded that standardized tests have had a somewhat detrimental to very detrimental affect education in the last twenty years. Here are a few statements by interviewed teachers describing the historical affect of standardized testing on education in the last 20 years:<sup>198</sup>

---

quoted words are not word for word perfect. They are transcribed notes with the intent to match the wording as close as possible.

<sup>197</sup> This rubric is apart of the Teacher Advancement Program Leadership Handbook, (193-200). The rubric was originally attached to an email from the superintendent, October 1, 2010. I cannot find the rubric on any TAP website or the district website page.

<sup>198</sup> The oral history given by these teachers for this particular historical question merits the large space devoted to their statements.

Well to be perfectly blunt--it has taken over education. I'm not certain about the last 20, but certainly the last 10-15 years. Now, there are consequences if you don't perform well on the test, so being human nature, teachers will teach towards that, and administrators will provide the culture of the school for that. Most of the resources today are for that very specific goal. It is just a fact of life. You have to make your peace with that, or you just throw up your hands and say I can't do anything. I have no problem with that as long as the tests are used properly, and that is just one little bit of data, but that doesn't seem to be the case. That is what is most frustrating to teachers, and the way politicians and the media present it. Most teachers know that there is so much more to education than that, because of that, it can be quite depressing to think about, and it seems to be getting worse year after year after year--its all data--its not the student at all. I find that, and most teachers I know, find that really offensive.

---

Standardized tests have totally shaped education in the last 20 years thanks to NCLB. Everyone is under the gun, if we don't get a certain score, than we are given sanctions, and that's crazy, I think teachers and administrators know that project-based learning, and authentic assessment, and writing, is the way to go. Why do they use standardized tests? It is cheaper. I went down once to grade this biology assessment, they were desperate, they took language teachers to grade this portfolio--this type of assessment failed miserably because it cost too much money.

---

Standardized tests have shaped education pretty darn huge in the last 20 years. The pendulum has swung way over. Maybe it will come back now. I have noticed that in the last 15 years it has been teach-to-the-test. I think the whole idea behind NCLB and 2014, they are realizing that these types of goals are not achievable, so they have to rethink what kinds of things they want to use. So I have noticed the pendulum swing back a little with using different types of assessments and strategies, and that's part of this whole new TAP thing that Sycamore River is doing, where we are doing an in-house thing looking at and working with our own. Localizing it.

---

In the early nineties, there was something called CAP and there were eight writing styles that they were supposed to write in. In the mid nineties, we had a really authentic measurement, I thought. It was called CLAS (California Learning Assessment System). The problem was conservative parents took issue with some of the questions that were asked in response to some of the literature the students read. *Has your family ever experienced something similar to this character?* They successfully squashed that test. I thought it was the only legitimate standardized test that I had seen, because it walked them through the story over the course of three or four days. They wrote an essay at the end, so it replicated what actually occurs in the classroom.

---

Standardized testing has shaped our education a lot. In first grade, we were able to do more one-on-one assessment when we were calling them back, and now



they are circling answers on a bubble sheet. The state has put a lot more weight into that test. We are looked at. We are compared, and we are really spending a lot of time getting our kids ready for a test rather than to be learners and to be excited about learning and to want to learn. Our classrooms are not as exciting as they used to be. It is not as fun. They don't get to do as much project-based learning where they are doing multiple things. So this week, after the test, everyone is quickly getting the projects out. Instead of all year long like it really should be. It is just part of what is happening, but it's just kind of where the educational system is right now.<sup>199</sup>

The TAP model uses standardized tests for half its evaluation of teachers. Why would teachers who feel this way about standardized tests support a system that uses standardized tests as the only measure of student achievement? In the interviews of teachers, two main reasons were given by next year's TAP teachers. First, NCLB consequences are very real and painful, and so any system that can help raise scores will help prevent the implementation of punitive actions. Second, they welcomed the teacher coaching and collaboration under the TAP system as a way to improve their teaching.

Survey Question Three stated, *Sycamore River's plan for implementing TAP teacher evaluations uses California Standardized Tests for half the teacher's evaluation. Consider how teaching for standardized tests will compare to the teaching of 21st Century Learning Skills. In Sycamore River, in a TAP classroom--* followed by three options: Teaching for Standardized Testing will be *greater than* (70), *less than* (36), or *equal with* (36) the teaching of 21 Century Learning Skills. Almost half the teachers surveyed predicted that, under the TAP model, teaching for the outcome of standardized tests will be *greater than* teaching for the outcome of 21st century learning skills. Only one-fourth of the teachers surveyed felt that standardized tests will be *less than* teaching for the outcome of 21st century learning skills.

---

<sup>199</sup> Four of the five teachers quoted here will be at a TAP school site next year.

Question Three showed that 75% of the teachers believe that teaching for the outcome of standardized tests will be equal with or greater than the teaching for the outcome of 21st century learning skills. Question Two showed that 78% of the surveyed teachers felt that standardized tests had a negative affect education in the last twenty years. The survey results from Question Two and Question Three help illustrate the disconnect found between educators and policymakers. It seems that the major reason there is any support among educators for standardized testing in its present form is to avoid NCLB sanctions, not because educators see value in standardized tests. TAP is designed to encourage better teaching. The student achievement growth measured by TAP does not have to be standardized tests. Sycamore River could choose its own assessment tools to use besides the California Standards Tests.

Fortunately, the Sycamore River Unified Teachers Association negotiated a one year agreement with the district regarding the implementation of the TAP system in Sycamore River. After one year of its implementation, teachers will have an opportunity to vote to continue or not continue TAP in its present form. I expressed concern over the use of standardized tests with a couple of the members of the district's teachers' association executive committee and they both stated that standardized tests are the easiest way to assess the students, since it is a measurement that is already in place, but with more time maybe authentic assessments could replace the standardized test results that are used as half the teacher's evaluation for a bonus.

Within the original vote to ratify TAP, 81 members voted yes, and 50 voted no. This vote is close when considering the fact that all the teachers that will be at a TAP site next year (2011/2012) will receive a \$3,000 bonus just for being at the TAP site, so many

of these voters had a vested interest in voting yes. Any teacher not at a TAP site technically had nothing to lose, because the TAP requirements only pertain to teachers at TAP sites. The money to implement TAP in Sycamore River will only come from Federal Title 1 and Title 2 funding and from the TIF grant. It is promised to not come from the general budget. Only 131 teachers voted. At the time of the vote, only seven schools out of the district's 17 schools would be TAP sites the following year. I asked several teachers who I know are against TAP, why they didn't show up to vote. Their response was that it didn't affect them, so they didn't care.

That attitude misses the fact that Sycamore River is being carefully watched as a test site to see how an alternative form of teacher evaluation could be used in California. An effort exists to change teacher tenure and make it more performance-based with the "performance" connected to standardized test results. It was reported that Sycamore River is a "crack in the wall of resistance to overhauling how teachers are evaluated and rewarded for their performance."<sup>200</sup> The report went on to explain that a California State Senator sponsored a bill to change teacher evaluations in California. The bill, SB 355 would have allowed multiple measures to be used in the evaluation as long as standardized test results were at least 30 percent of the evaluation. Districts could lay off teachers based on the performance of their students' test scores. Fortunately, this bill failed, but it was among several considered in Sacramento that challenge the current

---

<sup>200</sup> John Fensterwald, "Experiments in evaluating teachers: Districts and charters breaking new ground," Thoughts On Public Education, Simi Valley Education Foundation (June 1, 2011), <http://toped.svefoundation.org/2011/06/01/experiments-in-evaluating-teachers/> (accessed June 14, 2011).

teacher evaluation system. A number of programs and teachers around the state are piloting new evaluation methods.<sup>201</sup>

Many of the teachers that I interviewed do have a problem with the current evaluation system, but none of them felt that standardized test results should be used to lay off a teacher. Here is what some of the teachers said:

I think that standardized tests and teacher merit pay are not good ideas. I don't think standardized tests are any way to measure a teacher's performance. Merit pay may be simply rewarding teachers that are in an ideal situation and punishing teachers that are stuck in a bad situation.

I have no problem with teacher merit pay, but to start with standardized tests, that are grotesquely flawed in almost every category, that you would look at a teachers evaluation based on that, then no one has shown how you can have merit pay that is not flawed, for the same reasons that standardized tests are flawed. There are political aspects of evaluations. Also, what if I have special needs students. There's over and over and over a flawed system. There would have to be a very complicated system that uses many measures to see if a teacher is doing his job, and even then it could be politicized. There are some teachers that should not be teaching, there should be an easier way to get rid of those teachers.

Standardized testing and teacher merit pay--we could have a bonus system for teachers that are going above and beyond. In our current system, we are suppressing the incentive to work hard. If we could have a better assessment system of kids, then we could give rewards to teachers that go above and beyond. Project-based learning and authentic assessments can be connected to teacher bonuses. We could reward teachers who want to "move."

From my experience, teachers do not collaborate enough about issues of teacher merit pay and standardized testing. Instead, it seems they assume their union representatives will deal with it. The ratification vote to allow TAP in Sycamore River illustrated this lack of communication between teachers. Open communication was challenged in several ways on the day of the general membership vote to ratify TAP,

---

<sup>201</sup> John Fensterwald, "Experiments in evaluating teachers: Districts and charters breaking new ground," Thoughts On Public Education, Simi Valley Education Foundation (June 1, 2011), <http://toped.svefoundation.org/2011/06/01/experiments-in-evaluating-teachers/> (accessed June 14, 2011).

January 26, 2011. First, the vote in January occurred before the general membership knew enough details about TAP. The discussion about TAP barely started in September when most teachers were consumed by starting their school year. The information from the union president in September and October expressed concern over the TAP program. In a memorandum dated October 6, 2010, the union president explained frustration that the district had not sat down to negotiate the terms of TAP at that point. He wrote, “Our [name deleted] Negotiations team has been willing and ready to bargain the effects of the grant since last June. We urge the District to join us at the table as soon as possible.”<sup>202</sup> In November, the union president stated that we would be discussing this matter in more detail. In education, matters dealing with anything outside the classroom almost do not exist during the month of December for teachers, because of the ending of a semester, the closing-out of grades, and the approaching winter break and holidays. So, the next month that really counted after November, after being told we would discuss the matter more, was January. Instead of discussing the matter more, we voted on the ratification of TAP. Maybe the concerns that the union leadership held at the beginning of the year were put at ease when the union leadership realized that TAP evaluations would only be used for bonus purposes and would have nothing to do with job security. Whether they felt that way or not, the rest of the general membership needed an opportunity to meet and discuss the issues surrounding TAP, because as stated previously, a more ominous agenda exists that threatens teacher job security based on student standardized test results, and the adoption of TAP in Sycamore River is considered a crack in the wall of resistance that California has successfully maintained until now. Many Sycamore River teachers I spoke to were not aware of this. Sycamore River Unified School District covers 550

---

<sup>202</sup> Memorandum from the union president, October 6, 2010.

square miles. Rarely, teachers have an opportunity to come together and discuss important matters. With a district adoption of something as significant as TAP, the general membership is normally given a chance in an open forum to discuss the ramifications of such an adoption, before a vote is taken.

The Sycamore River TAP system will set an important precedent in California by permitting teacher evaluations to be determined by standardized test results. Proponents of the Sycamore River TAP system argue that it is an evaluation that is only used for a bonus. It is naive to think that this precedent setting move by Sycamore River will not feed into the contemporary argument by some groups that teacher job security should also be connected to standardized test results. Miscommunication and a hurried pace helped usher TAP into Sycamore River. In closed session, January 26, 2011, before the general membership was allowed to vote, the union representatives met for a regularly scheduled meeting. The issue of counting absentee ballots came up. Several of the union representatives collected ballots of members who knew they wouldn't be able to make it to the voting location during the open voting period. A debate ensued over whether absentee ballots should be counted or not. Ultimately, it was decided that absentee ballots would not be allowed, which was a controversial decision alone, because they have been allowed in the past. This debate caused the closed session to go longer than scheduled. Meanwhile, members waited outside for the doors to open so they could cast their vote. Instead of a devoted time for general comments regarding the issues surrounding TAP, before the vote occurred, the members came in and many of them began voting, and the union executive leadership did nothing to stop the vote to allow a discussion. While lines formed for members to vote, other members who wanted to

discuss the issues surrounding TAP sat in chairs hoping for a discussion of some kind. Members were allowed to write a comment or a question on a piece of paper, then one of the union executive members read the question or comment and then gave his response. This attempt at a discussion became a one-way conversation controlled by the union leadership and drowned-out by the commotion of other members talking while they waited in line to vote. I commented to the union treasurer about how we needed to discuss the fact that we do *not* have to use standardized test results in the TAP system. He thought standardized tests results were a required aspect of TAP. It is hoped that when the next vote is conducted in one year, the push to connect test scores to job security will be better understood by teachers, so that they can be better informed about the issues for which they vote.

Survey Question Four showed that teachers already understand how much a student's performance is affected by forces outside the control of the teacher or school. It will remain to be seen how much teachers get frustrated over an evaluation system that relies so heavily on these test results. Survey Question Four stated, *How much is a student's performance on standardized tests influenced by outside forces--either helped or hurt by things outside the realm of the teacher or school?---*followed by three options: *Very Influenced (97), Somewhat influenced (43), Not at all influenced (2)*. Half the overall evaluation of a Sycamore River TAP teacher is based on standardized tests, but 98.5% of the surveyed teachers believe that a student's performance on standardized tests is somewhat influenced to very influenced by outside forces.

Why would teachers support an evaluation system that is not completely within the control of the teacher? Again, a major reason stated by teachers is that NCLB pays

the bills, so it is beneficial if TAP helps to raise test scores. Here are the comments from several of the interviewed teachers:

Connecting standardized testing and teacher merit pay opens up a whole can of worms that generally is not good. I think that most professionals know the right thing. One of our goals in the last few years is to try to be everything for everyone. We are trying to do it all. We're trying to do the right thing, but at the same time produce the results with the data, so that we don't get into any kind of trouble or get into Program Improvement, because when that happens it takes so much of our resources away from what we feel is important.

---

Unfortunately, standardized tests are tied to funding, so they become a big deal, and it has shaped education. [name of school deleted] being a Program Improvement school--it is not fair, they are not measured by what they are doing.

---

Why do some advocate standardized testing--it gets the bills paid. If we're not doing well, then we won't be able to hire and we won't be able to provide for students. I think it needs to be corrected. I think something needs to be looked at that can bridge the gap between standardized testing and project-based learning. There's a point to standardized testing that is important, but so is authentic assessment. There needs to be a happy medium between the two.

---

It is tied to funding. The way the schools are perceived in the community--real estate, schools in the community, ADA. The shifting--public schools that have lost funding to charter schools. It is a resegregation. Well-to-do whites are moving into charter schools. A downward cycle begins and leads to ramifications. Bad scores for a few years can lead to well-to-do students leaving, which only makes future years even harder to raise scores.

---

I don't know a single educator or administrator who doesn't question the ridiculousness of the NCLB requirement, it is a standard that is ridiculously unattainable, and it is a disservice, because it just puts everyone on a grading scale--an A--you better get 99%. We have a school, [name deleted] that is at 850 or above API that is on Program Improvement. It is because they need to work with their ELL students more effectively, but you can't cut funding for a school like [name deleted]. The public is hyper-focusing on testing, but there are a whole bunch of other things that we are not paying attention to, all these kids that we are servicing. We need accountability and we can have a reward system, but it needs to be more equitable, more real, and truly more accountable. Real accountability to me is--does a kid walking out the doors of this school have the skills to function in the world, to participate in the democracy, to get a job, to make a living in life, and to understand enough to be able to access the information out there and use it appropriately, and that is the true test of whether we have done our job.



Survey Question Five stated, *Sycamore River should place more emphasis on teaching for Standardized Testing*--followed by four options: *Strongly Disagree* (80), *Somewhat Disagree* (45), *Somewhat Agree* (13), and *Strongly Agree* (4). According to 88% of teachers surveyed, Sycamore River should *not* place more emphasis on standardized testing, and yet to do well on the evaluation for TAP, teachers will need to raise their student scores from the previous year. TAP will be at six Sycamore River schools next year, four elementary schools, and two middle schools. So to obtain a TAP bonus, teachers at six schools will be placing more emphasis on standardized testing.

Survey Question Six stated, *Sycamore River should place more emphasis on teaching 21st Century Learning Skills*--followed by four options: *Strongly Disagree* (4), *Somewhat Disagree* (5), *Somewhat Agree* (71), and *Strongly Agree* (62). Ninety-four percent of teachers somewhat agreed to strongly agreed that Sycamore River should place more emphasis on 21st century learning skills. One of the interview questions touched on Question Five and Six of the Survey. It was a two part question that stated, *What type of learning outcomes are standardized tests good at assessing? What types of learning outcomes are not assessed by standardized tests?* Here are statements from several of the interviewed teachers regarding these questions:

Standardize tests are good at explaining demographic groups and student populations and what is going on there. If you have a student that is taking the test seriously, then maybe that is useful, but overall it is a forced assessment that isn't a valuable tool. There are a lot of things that are not assessed by standardized tests like life skills, things that students need to learn, we need to do more of teaching those things, life after school, and what it might be like if you don't have a bachelor's degree or a master's degree.

---

If the test is written correctly, it should correlate with the standards that should be taught in the class. Sometimes the tests don't match the state standards. What is not assessed---21st century skills are lacking--inquiry, questioning, higher level thinking, writing process, standardized testing is only doing paper to pencil--it is

not testing them orally, visually, auditor-ally. There are things that are left out of standardized testing.

---

Standardized tests are good at assessing--probably math. Math lends itself to answers, either right or wrong. Even math text books, the problems we have, could be multiple choice. Other things that come along are more difficult. Writing is more subjective. It works through the thinking process. Reading can be assessed by standardized tests, but to really get at how they are thinking or interpreting things, standardized tests can give a fair reading of where they are at, a snapshot for the moment.

---

What are not assessed by standardized tests--the thinking process, how are they going about attacking problems. Also, work-ethic is not assessed. A lot of students may not go onto college or higher education in general. So, really somebody that may kill it on a standardized test, may have the worst work-ethic. Maybe he doesn't turn in his work. That is not going to serve him well in the work environment. You could be a genius, but that doesn't mean anything. Standardized tests don't measure how they are going to do later on.

Unfortunately, Survey Question Seven cannot be counted because of the options provided. The question was *Does teaching for Standardized Tests and teaching 21st Century Learning Skills compliment or contradict each other*. The four options provided were: Strongly contradicts, Somewhat contradicts, Somewhat compliments, Strongly contradicts. I accidentally wrote strongly contradicts twice, instead of providing the option of strongly compliments. Ironically, this points to the frustration of standardized questions. The survey would have been much better if it just allowed written comments for each question. However, 142 survey responses would have presented a challenge to identify any patterns in the open-ended responses. This is where proponents of standardized tests have the strongest argument. Standardized test results provide the easiest way to compare students, but the price for comparisons are simplification and limitation. To overcome the challenge of limited options in the survey, optional comments were allowed. Those comments are included in Appendix B. Only a few comments were negative, regarding the survey. A few teachers expressed frustration

over the limited options for each question. That frustration speaks volumes about the whole issue of standardized testing versus authentic assessment where learning can be demonstrated in a more open-ended way. One teacher complained that the options didn't allow for any neutrality. The survey did force teachers to choose one side or the other, but isn't that action forced upon all of us. I am not sure there exists a luxury of neutrality regarding these issues. The stakes are too high.

Survey Question Eight stated, *What do you think are the findings of a majority of non-profit research experts REGARDING Multiple Choice Standardized Tests*--followed by three options: *Are effective tools for assessing learning (4)*; *Are not perfect, but are necessary for comparisons (56)*; *Are not accurate, and narrow the curriculum (82)*.

Fifty-eight percent of teachers recognized the same findings of this paper, that a majority of the research experts are against standardized testing. Many of the interviewed teachers felt that you could find any expert to say anything depending on what bias drove his or her motives. That is a correct observation. There certainly are experts devoted to both sides of the issue, and it often comes down to ulterior motives. The problem with that notion, though, is that it shows that teachers do not know that an overwhelming amount of research findings argue against standardized testing. Perhaps, if more teachers understood this, then maybe that knowledge would empower more teachers to speak out against standardized testing. Here are some of the responses by interviewed teachers:

---

Research experts say an ineffective tool, probably, an ineffective tool for getting an accurate measurement.

---

Experts--depends on their agenda or their bias.

---

The experts think it is a good idea.

---

Research experts know that standardized testing have failed.

---

Research experts are thinking that standardized testing is a good way to go, but I am feeling that a lot of people are realizing that there are others ways to go. Other schools are having tremendous success with project-based learning and other authentic measures.

---

Educational experts--depending on the bias or agenda for how they are approaching the data, will determine the outcome.

---

Research experts say that it provides some kind of norm or baseline in order to determine if students are grasping knowledge or not. That would be my biggest guess.

---

Research experts---I don't know that I have read that much about it. The stuff I have read is more through the teacher's unions and that tends to be more biased against merit pay. I have read that is doesn't necessarily increase student performance. Doesn't seem to be a real high correlation between the two.

---

Research experts--I know there is, I am sure there is research on both sides--just like a drug company--they can get the right pool of people so they can get the results they want to show its effective or not effective. The same is true of teacher merit pay systems. There are research experts on both sides of that. It might be true that if we focus on the tests, that we will get our test scores up. It might be true that kids are doing better in math and language arts, but we have left out this whole other part of what it means to be a well rounded educated person. I think that weighs just as much as having strong reading skills and math skills. Of course those two are primary and critical in order to access all the other subject areas, so as far as the research, I am sure it covers both sides of that.

Survey Question Nine stated, *If NCLB requirements were dissolved, would that change your opinion on these responses?*--followed by *Yes (50) or No (92)*. I included this question in my survey, because of input I received in the interviews. Teachers told me that, like it or not, the consequences of NCLB are very real, and so the test results must go up each year to avoid Program Improvement. Again, a common response in the interviews was that the test results pay the bills. This is how the teacher from the neighboring district described it:

In our district, unfortunately, the pressure to do well on standardized tests has increased, because we have been in Program Improvement for, I think three years now. And the unfairness of it is, my understanding at our school is, only fifty out

of two thousand students have kept us in Program Improvement, because they are a part of a certain subgroup of students that didn't do well enough on, not the CST, but on the CAHSEE. We have had an 85% pass rate. If you had any other test in the world where 85% of your population is passing, and you were that successful, hey your school deserves a solid "B," but instead, either you have an "A+" or you have an "F" when it comes to No Child Left Behind. We are a school that needs improvement, every school does, but we are not a Program Improvement school. I take issue with how we are being dubbed.

Teachers that I interviewed feel the squeeze of policy set down by NCLB. Until teachers across the nation unite in larger numbers and actively push back against the standardized testing movement, until there exists a viable agency to fight against NCLB, an individual teacher can only play the game and try to keep test scores up enough to help his or her school stay out of Program Improvement, while attempting to give students opportunities to learn the skills viable for this present age. It is disappointing to me that groups such as TAP and New Tech High, who have the ear of top policymakers and the media, do not push to dramatically shift the educational emphasis away from standardized testing. Maybe they are attempting to do this, but nothing tangible is occurring at the teachers level, and nothing is visible in public communiques. Maybe Congress is in shock and is frozen with inaction, because their bluff of NCLB-take-over power is about to be revealed as an over-inflated threat that cannot be carried out. Admitting NCLB was a mistake would be an invitation from Congress to discuss the legitimacy of the Federal Government's power-hold over education.

Secretary of Education Arne Duncan predicts that 80% of public schools will be deemed *failing* by 2014 under NCLB sanctions. He is considering allowing waivers to schools who embrace educational reform that President Obama has endorsed. One of those reform plans by President Obama is the Race-to-the-Top legislation that provides

money for the TIF grant.<sup>203</sup> Sycamore River is one of only 62 applicants in 27 states that received the TIF grant. If Secretary Duncan does allow for waivers to the 2014 NCLB deadline, Sycamore River will be strategically positioned to receive a waiver.

The leadership of Sycamore River's superintendent and its school board should be commended for embracing reform. The superintendent described this type of leadership displayed by the school board with their decision to implement New Tech High and The International Baccalaureate (IB) program:

Rather than circling the wagons and battening down the hatches, our board has chosen the bold alternative of rethinking the status quo and moving ahead with these innovative programs to ensure our students are competitive in the world economy. Their decision to move forward shows great leadership in the face of uncertainty and adversity, and gives us hope and optimism that we can meet the tough challenges ahead.<sup>204</sup>

Sadly, it doesn't appear that the elevated status of standardized testing is going anywhere, even in the midst of reform to encourage more teaching of 21st century learning skills. Secretary of Education Arne Duncan and President Obama are proponents of teacher merit pay, even in the midst of evidence that shows that teacher bonuses don't raise student test scores.<sup>205</sup> The performance measurement used to determine merit pay will continue to be standardized test results, so the potential relaxing of NCLB requirements through waivers for schools will merely shift the burden of those test results from schools to individual teachers.

---

<sup>203</sup> Sam Dillon, "Education Secretary May Agree to Waivers on 'No Child' Law Requirements," *The New York Times*, June 12, 2011, [http://www.nytimes.com/2011/06/12/education/12educ.html?pagewanted=1&\\_r=1](http://www.nytimes.com/2011/06/12/education/12educ.html?pagewanted=1&_r=1) (accessed June 14, 2011).

<sup>204</sup> Superintendent's blog found at Sycamore River's Website under his entry on April 14, 2011 [district website deleted for anonymity reasons] (accessed June 15, 2011).

<sup>205</sup> Christopher Connell, The Hechinger Report "Merit pay study: Teacher bonuses don't raise student test scores" in USA Today, September 21, 2010, [http://www.usatoday.com/news/education/2010-09-21-merit-pay\\_N.htm](http://www.usatoday.com/news/education/2010-09-21-merit-pay_N.htm) (accessed June 17, 2011).

What remains to be seen is how much Sycamore River will continue to shape its instruction around standardized testing. According to the vision laid out by the superintendent, Sycamore River will de-emphasize the relative importance of standardized testing, so that it becomes a smaller part of the district's overall instructional goals. In the 21<sup>st</sup> Century Learning Community Forum, hosted by Sycamore River on March 30, 2010, the superintendent explained why standardized testing in Sycamore River needs to be relatively de-emphasized:

Currently, our results are primarily measured by state test scores in math and Language Arts, which is a very limited assessment of all we need to know about our students. **Our students need to master the basic skills, but they also need so much more:** they need to be equipped with an arsenal of more complex skills that will help them to navigate in a very challenging, competitive and technologically developed world. In addition to our standardized assessments, we will need to find more authentic ways of measuring the more complex skills that are needed today. **There is so much more to education than bubbling in a Scantron sheet.** [bold font apart of original text]<sup>206</sup>

To hear a superintendent in a forum about vision-casting make such a bold statement against standardized testing is a hopeful sign that things are starting to turn around.

Historically, it was not long ago that NCLB was born, but in only one decade, NCLB has grown into the sacred cow that represents federal power over education, and both Democrats and Republicans embrace it as such. The issues in this paper, in so many ways, deal with federal power over education. When the power is pulled away from local communities, then the only viable assessment tool within the power of the Federal Government is standardized testing. In a very small way, I learned about the convenience of standardized multiple choices from conducting my survey. Federal power over education is a vicious circle where local districts struggle to meet federal NCLB

---

<sup>206</sup> 21<sup>st</sup> Century Learning Community Forum: Introduction by the Sycamore River superintendent, March 30, 2010. [district website deleted for anonymity reasons] (accessed June 14, 2011).

requirements, and so they apply for federal grants to do whatever they can to meet those requirements, because if they don't meet those requirements they lose federal funding. In the midst of this vicious circle, local administrators make decisions that are contrary to quality education. One teacher I interviewed explained the trend in the late 1990s that elevated preparation for standardized tests above all other learning:

Standards changed in the late nineties. In 1999, we were issued a directive from our superintendent down to the mentors to focus on nothing but test prep, test prep, test prep, test prep. *All those other things you are doing, those were great, but we need to raise our test scores.* From that day forward it has been a downward slide toward the "right" answer, which is the exact opposite of what you want in the humanities. You want to provide the question, the text, and the possibility for many answers. As you know, a well written poem will have as many interpretations as there are readers in that classroom. Some readings are probably more accurate than others, but you want the connotations to take flight, and not narrow it down to, *no--the correct answer to this poem is...*

Sycamore River's Vision Statement in its 2020 Vision Framework is to "become the model school district for 21st Century Learning in the nation." These skills identified in this vision framework are: 1) Communication, 2) Problem Solving and Critical Thinking, 3) Teamwork, Collaboration and Cooperation, 4) Technology, 5) Self-Direction 6) Innovation, Imagination and Creativity 7) Global Awareness & Second Language.<sup>207</sup> Lately, Sycamore River has proven to be very effective at creating vision and planning programs, but the actual implementation of a new reality that is more focused on authentic assessment and less obsessed with simple multiple choice tests will be true evidence of leadership.

In his superintendent's blog, March 7, 2011, summarizing his experience at the CSU Summit about Transformative Change in the Preparation of Teachers, the superintendent expressed his excitement about how TAP will help Sycamore River close

---

<sup>207</sup> Sycamore River Unified School District, "Vision 2020 Framework," Sycamore River Unified School District. [district website deleted for anonymity reasons] (accessed November 23, 2010).



persistent achievement gaps in economically disadvantaged populations, students with special needs, and English learners. In the very next paragraph the superintendent described how impressed he was by one of the speakers Dr. Linda Darling-Hammond:

Most of the CSU presidents were there, along with deans and professors from various fields. The guest speakers and all the individual workshops were very informative. **One speaker stood out to me in particular, Dr. Linda Darling-Hammond**, who is an education professor at Stanford. Her resume and achievements are extremely impressive. She is a former president of the American Educational Research Association and a member of the National Academy of Education. Her research, teaching, and policy work focus on issues of school reform, teacher quality, and educational equity. From 1994-2001, she served as executive director of the National Commission on Teaching and America's Future, a blue-ribbon panel whose 1996 report, "What Matters Most: Teaching for America's Future," led to sweeping policy changes affecting teaching in the United States. In 2006, this report was named one of the most influential affecting U.S. education, and Darling-Hammond was named one of the nation's 10 most influential people affecting educational policy over the last decade. In 2008-09, she headed President Barack Obama's education policy transition team. Her comments on the positive changes taking place in our teacher education programs, and what needs to continue to change, were very enlightening. It was great to hear her thoughts and ideas, considering her extensive background and comprehensive work with education reform.<sup>208</sup>

The reason I presented so much of the superintendent's wording here is to show the disconnect that occurs. One paragraph shows excitement about TAP, and the very next paragraph describes in great detail how impressed he is with Dr. Linda Darling-Hammond and all the expertise that she represents. Remember, Linda Darling-Hammond is one of the ten experts from the Economic Policy Institute that wrote a report warning against value-added models, the same methodology that TAP uses. That report, discussed in the previous section of this paper, also warned against standardized testing, the tool of measurement that Sycamore River chose to use in its TAP system. Again, maybe the superintendent realizes this disconnect exists, and maybe he plans to use TAP

---

<sup>208</sup> The superintendent blog can be found on the district website, [district website deleted for anonymity reasons] (accessed June 14, 2011).

simply as a system to raise test scores to avoid NCLB sanctions. One hopeful sign in this scenery is the fact that Linda Darling-Hammond, an educator, and a research expert with findings that helped shape this paper, advised President Obama about educational policy.

I want to believe that the superintendent's vision will become a reality, where Sycamore River will focus on 21st century learning skills through more project-based learning while de-emphasizing instruction related to standardized testing. I want to believe that Sycamore River will shift its TAP student achievement measurement from standardized test results to authentic assessment after NCLB restrictions over schools relax. I wonder how long TAP will remain after the TIF grant runs out in five years. I wonder if TAP will morph into something more ominous, away from bonus pay to job security.

Currently, the stakes are ratcheting-up in the nation, beyond the main subject of this paper, *standardized testing and teacher merit pay*, to *standardized testing and the job security of individual teachers*. Evidence shows that teachers don't respond to bonuses like lab rats to food.<sup>209</sup> As policymakers realize this, their temptation will be to tie the incentive, not to bonuses, but to job security. Lately, in New York, Los Angeles, and Seattle, to name a few places nationally, this idea of using test results to evaluate teachers for the consideration of teacher layoff is gaining traction.<sup>210</sup> Linda Darling-Hammond just wrote a piece in *The New York Times* challenging the idea that more tests (the results from a new battery of tests for New York students that will exist solely to evaluate the

---

<sup>209</sup> Christopher Connell, The Hechinger Report "Merit pay study: Teacher bonuses don't raise student test scores" in USA Today, September 21, 2010, [http://www.usatoday.com/news/education/2010-09-21-merit-pay\\_N.htm](http://www.usatoday.com/news/education/2010-09-21-merit-pay_N.htm) (accessed June 17, 2011).

<sup>210</sup> Author unknown, "All the Reasons Why Teacher's Evaluations Should Not Be Tied to Test Scores," Seattle Education: News and Commentary, January 8, 2011, <http://seattleducation2010.wordpress.com/2011/01/08/all-the-reasons-why-teachers-evaluations-should-not-be-tied-to-test-scores/> (accessed June 17, 2011).

teacher for job security purposes) is not what American students need, because the tests pull precious time away from critically needed instruction. She wrote, “Recent research shows that test scores are highly unstable and error prone for measuring individual teachers, and that making high-stakes decisions based on these tests causes schools to reduce their teaching of important content and skills not measured by the tests.”<sup>211</sup>

Lately, the frustration by some Sycamore River teachers is that testing is taking over, and they are not just referring to the California Standards Test that is administered in the spring. Sycamore River now tests students regularly with pre and post tests to assess benchmark learning outcomes, and now students in Sycamore River increasingly face testing. Many teachers expressed related concerns in the optional comments section of my survey. It is hoped that these concerns will grow into a resistance against any use of standardized test results to evaluate teachers. I hope that the next time teachers vote for TAP, they realize how this system is being viewed by some policymakers as a way to shift the burden of NCLB from funding for schools to the job security of individual teachers. I hope more teachers share with each other what they shared with me. Here are concluding thoughts from those teachers:

Standardized testing is making our students LESS competitive in the global marketplace. We are not adequately preparing our students for the real 21st century world.

Which is more important, test scores or students being able to think outside the box to solve a problem? As long as test scores drive our schools, independent and group creativity in problem solving will slowly fall to the way-side. And that will be very sad.

---

<sup>211</sup> Linda Darling-Hammond, “A Dangerous Obsession (a response in Room for Debate: A Running Commentary on the News--Testing Students to Grade Teachers,” The New York Times, May 30, 2011, <http://www.nytimes.com/roomfordebate/2011/05/30/testing-students-to-grade-teachers/excessive-testing-is-a-dangerous-obsession> (accessed June 17, 2011)

I don't believe TAP will limit teachers teaching 21st century skills. I hope TAP will give teachers new insight on teaching strategies and help teachers engage students and motivate students to want to learn. Becoming a better teacher should be the focus of TAP, not teaching to a test. If teachers are doing a better job engaging and motivating students they should learn more and do better on valid assessments.

---

Teaching "to the test" should be outlawed. It has caused real learning to decrease because of the emphasis on having children memorize minutia at its worst and learning test taking strategies at the best. What a waste of time and money and energy for everyone involved.<sup>212</sup>

“Bring the past only if you are going to build from it.”<sup>213</sup> I believe that history is only meaningful when it possesses the power to move people today to make a change for tomorrow. From the historical vista of this paper the reader views two paths that join and form a trail of pitfalls. The use of standardized tests to evaluate the performance of teachers will block more meaningful learning. When individual teachers join to collectively share with society their lessons of experience with standardized testing, then maybe the media will focus on the findings of the educational research. When policymakers witness a mass-movement away from standardized testing, then lasting, effective, meaningful educational reforms will shape instruction and assessment that etch into the lives of our students skills that form the structural marrow of tomorrow.

---

<sup>212</sup> These comments are found in Appendix B.

<sup>213</sup> Domenico Cieri Estrada-unknown where this phrase first appeared.

## Bibliography

- Amrein-Beardsley, Audrey. "Value-Added Tests: Buyer, Be Aware." *Educational Leadership* 67, no. 3 (November, 2009): 38-47.
- Asia Society (author unknown). "What Accounts for Finland's High Student Achievement Rate?" *Asia Society*, April 27, 2010.  
<http://asiasociety.org/education-learning/learning-world/what-accounts-finlands-high-student-achievement-rate> (accessed June 25, 2011).
- Bacharach, Samuel B., Sharon C. Conley, and Joseph B. Shedd. "Evaluating Teachers for Career Awards and Merit Pay." In *The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers*, edited by Jason Millman, and Linda Darling-Hammond, 133-146. Newbury Park: Sage Publications, Inc., 1990.
- Bacharach, Samuel B., David B. Lipsky, and Joseph B. Shedd. *Paying For Better Teaching: Merit Pay and Its Alternatives*. Ithaca: Organizational Analysis and Practice, Inc., 1984.
- Baker, Eva L., Paul E. Barton, Linda Darling-Hammond, Edward Haertel, Helen F. Ladd, Robert L. Linn, Diane Ravitch, Richard Rothstein, Richard J. Shavelson, and Lorrie A. Shepard. "Problems with the Use of Student Test Scores to Evaluate Teachers." *Economic Policy Institute, Briefing Paper #278* (August 29, 2010). [http://epi.3cdn.net/724cd9a1eb91c40ff0\\_hwm6iij90.p](http://epi.3cdn.net/724cd9a1eb91c40ff0_hwm6iij90.p) (accessed October 20, 2010).
- Baker, Eva L. *Understanding Educational Quality: Where Validity Meets Technology*. Princeton: Educational Testing Service, 2000.
- Barton, Paul E. *A Policy Information Perspective: Facing the Hard Facts in Education Reform*. Princeton: Educational Testing Service, 2001.
- Barton, Paul E. *A Policy Information Report--Order in the Classroom: Violence, Discipline, and Student Achievement*. Princeton: Educational Testing Service, 1998.
- Barton, Paul E. *A Policy Information Report--What Jobs Require: Literacy, Education, and Training, 1940-2006*. Princeton: Educational Testing Service, 2000.
- Belanoff, Pat, and Marcia Dickson, eds. *Portfolios: Process and Product*. Portsmouth: Boynton/Cook Publishers, Inc., 1991.
- Bracey, Gerald W. "Big Tests: What Ends Do They Serve?" *Educational Leadership* 67, no. 3 (November, 2009): 32-37.

- Bracey, Gerald W. "Serious Questions About the Tennessee Value-Added Assessment System." *Phi Delta Kappan* 85, no. 5 (January, 2004): 716-717.
- Brandt, Richard M. *Incentive Pay and Career Ladders for Today's Teachers: A Study of Current Programs and Practices*. Albany: State University of New York Press, 1990.
- Braun, Henry, Naomi Chudowsky, and Judith Koenig, eds. *Getting Value Out of Value-Added: A Report of a Workshop, Committee on Value-Added Methodology for Instructional Improvement, Program Evaluation, and Educational Accountability*. Washington, D.C.: The National Academies Press, 2010.
- Brennan, Robert L., ed. *Educational Measurement: Sponsored Jointly by National Council on Measurement in Education and American Council on Education*. Westport: Praeger Publishers, 2006.
- Brookhart, Susan M. "The Many Meanings of 'Multiple Measures.'" *Educational Leadership* 67, no. 3 (November, 2009): 6-13.
- Buck Institute for Education: <http://www.bie.org/> (accessed January 20, 2011).
- California Department of Education. "California Standards Test-Released Test Questions" pages 13-14 (March, 2009).  
<http://www.cde.ca.gov/ta/tg/sr/documents/cstrtqhssmar18.pdf> (accessed November 23, 2010).
- California Educator*: California Teachers Association, 15, no. 5 (February 2011): 18.
- California State Board of Education-English-Language Arts Content Standards for California Public Schools-Kindergarten Through Grade Twelve (December, 1997). <http://www.cde.ca.gov/be/st/ss/documents/elacontentstnds.pdf> (accessed November 24, 2010)
- Clotfelter, Charles T., Helen F. Ladd, Jacob L. Vigdor, and Roger Aliaga Diaz. "Do School Accountability Systems Make It More Difficult for Low-Performing Schools to Attract and Retain High Quality Teachers?" *Journal of Policy Analysis and Management* 23, no. 2 (Spring 2004).  
<http://www.jstor.org/stable/3326150> (accessed October 15, 2010).
- Cohen, David K., and Marvin Lazerson. "Education and the Corporate Order." *In Education in American History: Readings on the Social Issues*, edited by Michael B. Katz, 318-333. New York: Praeger Publishers, 1973.

- Connell, Christopher. The Hechinger Report “Merit pay study: Teacher bonuses don’t raise student test scores.” USA Today, September 21, 2010.  
[http://www.usatoday.com/news/education/2010-09-21-merit-pay\\_N.htm](http://www.usatoday.com/news/education/2010-09-21-merit-pay_N.htm)  
 (accessed June 17, 2011).
- Cresap, McCormick and Paget (no other author information given). *Teacher Incentives: A Tool for Effective Management*. Reston: National Association of Secondary School Principals, National Association of Elementary School Principals, and American Association of School Administrators, 1984.
- Darling-Hammond, Linda. “A Dangerous Obsession (a response in Room for Debate: A Running Commentary on the News--Testing Students to Grade Teachers.” *The New York Times*, May 30, 2011.  
<http://www.nytimes.com/roomfordebate/2011/05/30/testing-students-to-grade-teachers/excessive-testing-is-a-dangerous-obsession> (accessed June 17, 2011)
- Darling-Hammond, Linda, Jacqueline Aneess, and Beverly Falk. *Authentic Assessment in Action: Studies of Schools and Students at Work*. New York: Teachers College Press, 1995.
- Darling-Hammond, Linda, Arthur E. Wise, and Stephen P. Klein. *A License to Teach: Building a Profession for 21st-Century Schools*. Boulder: Westview Press, 1995.
- Dillon, Sam. “Education Secretary May Agree to Waivers on ‘No Child’ Law Requirements.” *The New York Times*, June 12, 2011.  
[http://www.nytimes.com/2011/06/12/education/12educ.html?pagewanted=1&\\_r=1](http://www.nytimes.com/2011/06/12/education/12educ.html?pagewanted=1&_r=1) (accessed June 14, 2011).
- Dockery, Linda, and Marcia Epstein. “The Teacher Incentive Project (TIP) of the Winston-Salem/Forsyth County Schools.” In *Merit, Money and Teachers’ Careers: Studies on Merit Pay and Career Ladders for Teachers*, edited by Henry C. Johnson, Jr., 209-225. Lanham: University Press of America, Inc., 1985.
- Doran, Harold C. and Steven Fleischman. “Challenges of Value-Added Assessment.” *Educational Leadership* 63, no. 3 (November, 2005): 85-87.
- Dorn, Sherman. *Accountability Frankenstein: Understanding and Taming the Monster*. Charlotte: Information Age Publishing, 2007.
- Dwyer, Carol Anne, ed. *The Future of Assessment: Shaping Teaching and Learning*. New York: Lawrence Erlbaum Associates, Publishers, 2008.
- Estrada, Domenico Cieri-unknown where his famous quote first appeared--*Bring the past only if you are going to build from it*.

- Evans, Dennis L., ed. *Taking Sides: Clashing Views on Controversial Issues in Secondary Education*. Guilford: McGraw-Hill Companies, Inc., 2002.
- Evergreen Freedom Foundation. "Value Added Assessment." *Evergreen Freedom Foundation, School Director's Handbook* (n.d.) <http://www.myfreedomfoundation.com/pdfs/Value-Added.pdf> (accessed October 25, 2010).
- Fensterwald, John. "Experiments in evaluating teachers: Districts and charters breaking new ground." *Thoughts On Public Education*, Simi Valley Education Foundation (June 1, 2011). <http://toped.svefoundation.org/2011/06/01/experiments-in-evaluating-teachers/> (accessed June 14, 2011).
- Fisk, Edward B., and Helen F. Ladd. *When Schools Compete: A Cautionary Tale*. Washington D.C.: Brookings Institution Press, 2000.
- Foote, Chandra J., Paul J. Vermette, and Catherine F. Battaglia. *Constructivist Strategies: Meeting Standards and Engaging Adolescent Minds*. Larchmont: Eye on Education, 2001.
- Freeley, Mary Ellen, and Richard Hanzelka. "Getting Away from Seat Time: A New Hampshire initiative encourages schools to move toward competency-based learning." *Educational Leadership* 67, no. 3 (November, 2009): 63-67.
- Fuhrman, Susan H., and Betty Malen, eds. *The Politics of Curriculum and Testing*. Bristol: Falmer Press, 1991.
- Gipps, Caroline V. "Accountability Testing and the Implications for Teacher Professionalism." In *Measurement and Research in the Accountability Era*, edited by Carol Anne Dwyer, 99-112. Mahwah: Lawrence Erlbaum Associates, Publishers, 2005.
- Graham, Jared. "Finland test scores in perspective." *Education in Practice: Exploring Issues in Education and Education Reporting*. August 12, 2010. <http://www.educationinpractice.net/tag/international-test/> (accessed July 8, 2011).
- Gratz, Donald B. "The Problem with Performance Pay: It can work, but only if performance is broadly defined and all parties agree to the plan." *Educational Leadership* 67, no. 3 (November, 2009): 76-79.
- Haertel, Edward H. and Joan L. Herman, eds. *Uses and Misuses of Data for Educational Accountability and Improvement: 104th Yearbook of the National Society for the Study of Education, Part II*. Malden: Blackwell Publishing, 2005.



- Hawley, William. "The Limits and Potential of Performance-Based Pay as a Source of School Improvement." In *Merit, Money and Teachers' Careers: Studies on Merit Pay and Career Ladders for Teachers*, edited by Henry C. Johnson, Jr., 93-98. Lanham: University Press of America, Inc., 1985.
- Herndon, Terry. "Merit Pay and the Concerns of the Teaching Profession." In *Merit, Money and Teachers' Careers: Studies on Merit Pay and Career Ladders for Teachers*, edited by Henry C. Johnson, Jr., 93-98. Lanham: University Press of America, Inc., 1985.
- Holland, Paul W. "Assessing the Validity of Test Linking: What Has Happened Since Uncommon Measures?" In *Measurement and Research in the Accountability Era*, edited by Carol Anne Dwyer, 185-195. Mahwah: Lawrence Erlbaum Associates, Publishers, 2005.
- International Baccalaureate Program. <http://www.ibo.org/pyp/> (accessed June 18, 2011).
- Johnson, Henry C., ed. *Merit, Money and Teachers' Careers: Studies on Merit Pay and Career Ladders for Teachers*. Lanham: University Press of America, Inc., 1985.
- Jonassen, David H., and Susan M. Land. *Theoretical Foundations of Learning Environments*. Mahwah: Lawrence Erlbaum Associates, Publishers, 2000.
- Kane, Thomas J., and Douglas O. Staiger. "The Promise and Pitfalls of Using Imprecise School Accountability Measures." *The Journal of Economic Perspectives* 16, no. 4 (Autumn, 2002). <http://www.jstor.org/stable/3216916>. (accessed October 15, 2010).
- Kohn, Alfie. "Fighting the Tests: A Practical Guide to Rescuing Our Schools." <http://clogic.eserver.org/4-1/kohn.html> (accessed 10/31/2010).
- Kohn, Alfie. *The Case Against Standardized Testing: Raising the Scores, Ruining the Schools*. Portsmouth: Heinemann, 2000.
- Ladd, Helen F., ed. *Holding Schools Accountable: Performance-Based Reform in Education*. Washington D.C.: The Brookings Institution, 1996.
- Ladson-Billings, Gloria. "No Teacher Left Behind: Issues of Equity and Teacher Quality." In *Measurement and Research in the Accountability Era*, edited by Carol Anne Dwyer, 141-162. Mahwah: Lawrence Erlbaum Associates, Publishers, 2005.
- Larabee, Lottie, B. *Can Teachers Get Rich?* Garden City: Education Press, 1960.
- Levitt, Steven D., and Stephen J. Dubner. *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*. 3rd ed. New York, Harper Perennial, 2009.

- Linn, Robert L., and M. David Miller. *Measurement and Assessment in Teaching*. 9th ed. Upper Saddle River: Pearson Education, Inc., 2005.
- Loveless, Tom. "Test-Based Accountability: The Promise and the Perils." In *Brookings Papers on Education Policy*, edited by Diane Ravitch, 7-46. Washington D.C.: Brookings Institution Press, 2005.
- Manor New Tech High School. <http://www.manorisd.net/portal/newtech/> (accessed June 17, 2011).
- Meier, Deborah. *The Power of Their Ideas: Lessons for America from a Small School in Harlem*. Boston: Beacon Press, 2002.
- Meier, Deborah, and George Wood, eds. *Many Children Left Behind: How the No Child Left Behind Act Is Damaging Our Children and Our Schools*. Boston: Beacon Press, 2004.
- Moulthrop, Daniel, and Ninive Clements Calegari, and Dave Eggers. *Teachers Have It Easy: The Big Sacrifices and Small Salaries of America's Teachers*. New York: The New Press, 2005.
- Natale, Catherine Fisk, "Retaining and Rewarding the Best Teachers: 21st Strategies for Teacher Compensation: Models of Alternative Teacher Compensation Systems." CABE/CAPSS Convention (2009).  
[http://www.cabe.org/uploaded/Convention\\_Handouts/Performance\\_pay\\_models.pdf](http://www.cabe.org/uploaded/Convention_Handouts/Performance_pay_models.pdf) (accessed November 11, 2010).
- National Center for Education Statistics <http://nces.ed.gov/nationsreportcard/about/> (accessed June 24, 2011).
- National Center on Education and the Economy (U.S.). *Tough Choices or Tough Times: New Commission on the Skills of the American Workforce*. San Francisco: John Wiley and Sons, 2007.
- National Education Association. *SEARCH-Structuring Teacher Pay: Panaceas, Problems and Potential*. West Haven: National Education Association, 1985.
- NCLB Program Improvement School Requirements. <http://www.cde.ca.gov/ta/ac/ti/nclbpireq.asp> (accessed April 27, 2011).
- New Tech Network. <http://www.newtechnetwork.org/> (accessed June 17, 2011).
- Nichols, Sharon L., and David C. Berliner. *Collateral Damage: How High-Stakes Testing Corrupts America's Schools*. Cambridge: Harvard Education Press, 2007.

- Odden, Allan, and Caralyn Kelley. *Paying Teachers For What They Know and Do: New and Smarter Compensation Strategies To Improve Schools*. 2nd ed. Thousand Oaks: Corwin Press, Inc., 2002.
- Olson, Lynn. "Tennessee Reconsiders Value-Added Assessment System." *Education Week* 23, no. 25: 9 (2004) *Academic Search Elite*, EBSCOhost (accessed October 26, 2010).
- PACT (Performance Assessment for California Teachers) [http://www.pacttpa.org/\\_main/hub.php?pageName=Home](http://www.pacttpa.org/_main/hub.php?pageName=Home) (accessed November 1, 2010).
- Parker, James C. *Career Ladder/Master Teacher Programs: Implications for Principals*. Reston: The National Association of Secondary School Principals, 1985.
- Partnership for 21st Century Learning Skills. <http://www.p21.org/> (accessed November 23, 2010).
- Pearlman, Bob. "21st Century Learning in Schools--A Case Study of New Technology High School in Napa, CA." <http://www.bobpearlman.org/Articles/21stCenturyLearning.htm> (accessed May 12, 2011).
- Peterson, Paul E., ed. *Choice and Competition in American Education*. Lanham: Rowman & Littlefield Publishers, Inc., 2006.
- Phelps, Richard. "Why Testing Experts Hate Testing." In *Taking Sides: Clashing Views on Controversial Issues in Secondary Education*, edited by Dennis L. Evans, 156-168. Guilford: McGraw-Hill Companies, Inc., 2002.
- Ravitch, Diane, ed. *Brookings Papers on Education Policy*. Washington D.C.: Brookings Institution Press, 2005.
- Ravitch, Diane, ed. *Debating the Future of American Education: Do We Need National Standards and Assessments?* Washington D.C.: The Brookings Institute, 1995.
- Ravitch, Diane. *Left Back: A Century of Failed School Reforms*. New York: Simon and Schuster, 2000.
- Ravitch, Diane. *The Revisionists Revised: A Critique of the Radical Attack on the Schools*. New York: Basic Books, Inc., 1978.
- Ravitch, Diane. *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education*. New York: Basic Books, 2010.
- Rothstein, Richard. *What Do We Know About Declining (or Rising) Student Achievement?* Arlington: Educational Research Service, 1997.

- Rothstein, Richard, Rebecca Jacobsen, and Tamara Wilder. *Grading Education: Getting Accountability Right*. New York: Teachers College Press, 2008.
- Rumery, Robert E. "Measuring Teacher Merit: What We Can Learn from Businesses and Industry." In *Merit, Money and Teachers' Careers: Studies on Merit Pay and Career Ladders for Teachers*, edited by Henry C. Johnson, Jr., 139-167. Lanham: University Press of America, Inc., 1985.
- Ryan, Katherine E., and Lorrie A. Shepard, eds. *The Future of Test-Based Educational Accountability*. New York: Routledge, 2008.
- Sacks, Peter. "Standardized Testing: Meritocracy's Crooked Yardstick." In *Taking Sides: Clashing Views on Controversial Issues in Secondary Education*, edited by Dennis L. Evans, 169-175. Guilford: McGraw-Hill Companies, Inc., 2002.
- Sanders, William L. "Comparisons Among Various Educational Assessment Value-Added Models." SAS Institute, Inc. (October 16, 2006).  
[http://www.talentedteachers.com/pubs/comparisons\\_models\\_sanders.pdf](http://www.talentedteachers.com/pubs/comparisons_models_sanders.pdf)  
 (accessed November 11, 2010).
- Savage, Daniel M. *John Dewey's Liberalism: Individual, Community, and Self-Development*. Carbondale: Southern Illinois University Press, 2002.
- Schwartz, Daniel L., Robb Lindgren, and Sarah Lewis. "Constructivism in an Age of Non-Constructivist Assessments." In *Constructivist Instruction: Success or Failure*, edited by Sigmund Tobias and Thomas M. Duffy, 34-61. New York: Routledge, 2009.
- Seattle Education* (author unknown) "All the Reasons Why Teacher's Evaluations Should Not Be Tied to Test Scores." *Seattle Education: News and Commentary*, January 8, 2011, <http://seattleducation2010.wordpress.com/2011/01/08/all-the-reasons-why-teachers-evaluations-should-not-be-tied-to-test-scores/> (accessed June 17, 2011).
- Senechal, Diana. "The Most Daring Education Reform of All." *American Educator* (Spring 2010). <http://autoteacher.net/uploads/Senechal.pdf> (accessed October 22, 2010).
- Shavelson, Richard J. *A Brief History of Student Assessment: How We Got Where We Are and a Proposal for Where to Go Next*. Washington D.C.: Association of American Colleges and Universities, 2007.
- Solomon, Lewis C. "The Case for Merit Pay." In *Choice and Competition in American Education*, edited by Paul E. Peterson, 102-110. Lanham: Rowman & Littlefield Publishers, Inc., 2006.

- Solomon, Lewis C. and Michael Podgursky. "The Pros and Cons of Performance-Based Compensation." In *Taking Sides: Clashing Views on Controversial Issues in Secondary Education*, edited by Dennis L. Evans, 180-190. Guilford: McGraw-Hill Companies, Inc., 2002.
- Solomon, Pearl G. *The Curriculum Bridge: From Standards to Actual Classroom Practice*. Thousand Oaks: Corwin Press, 2009.
- Spring, Joel. *The Sorting Machine: National Educational Policy Since 1945*. New York: David McKay Company, Inc., 1976.
- Springer, Matthew G., Dale Ballou, Laura Hamilton, Vi-Nhuan Le, J.R. Lockwood, Daniel F. McCaffrey, Matthew Pepper, and Brian M. Stecher. "Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching." National Center on Performance Incentives, Project on Incentives in Teaching (September 21, 2010). <http://edlab.tc.columbia.edu/files/pointstudy.pdf> (accessed October 22, 2010).
- Springer, Matthew G., ed. *Performance Incentives: Their Growing Impact on American K-12 Education*. Washington, D.C.: Brookings Institution Press, 2009.
- Steffy, Betty E., and Fenwick W. English. *Curriculum and Assessment for World-Class Schools*. Lancaster: Technomic Publishing Company, Inc., 1997.
- Stone, John E. "Value-Added Assessment: An Accountability Revolution." Education Consumers Clearing House [http://www.education-consumers.com/articles/value\\_added\\_assessment.shtm](http://www.education-consumers.com/articles/value_added_assessment.shtm) (accessed April 20, 2011).
- TAP--The System for Teacher and Student Advancement-brochure [http://www.tapsystem.org/pubs/tap\\_brochure.pdf](http://www.tapsystem.org/pubs/tap_brochure.pdf) (accessed June 22, 2011). TAP is now run through the National Institute for Excellence in Teaching.
- Tanner, Daniel, and Laurel Tanner. *History of the School Curriculum*. New York: Macmillan Publishing Company, 1990.
- Tomlinson, Stephen. "Edward Lee Thorndike and John Dewey on the Science of Education." *Oxford Review of Education* 23, no. 3 (Sept. 1997), 365-383. <http://www.jstor.org/stable/1050962> (accessed September 6, 2009).
- Trilling, Bernie, and Charles Fadel. *21st Century Skills: Learning for Life In Our Times*. San Francisco: Jossey-Bass, 2009.

- Tucker, Bill. "The Next Generation of Testing: Simulation. Situated exercises. Tracking students' thought processes as they solve a problem. Welcome to the world of 21st-century assessment." *Educational Leadership* 67, no. 3 (November, 2009): 48-53.
- Urban, Wayne J. "Old Wine, Old Bottles?: Merit Pay and Organized Teachers." In *Merit, Money and Teachers' Careers: Studies on Merit Pay and Career Ladders for Teachers*, edited by Henry C. Johnson, Jr., 25-38. Lanham: University Press of America, Inc., 1985.
- Van Loozen, Lu. *Some Points to Consider When You Discuss Merit Pay*. Arlington: American Association of School Administrators, 1983.
- Wiley, Edward W. "A Practioner's Guide to Value Added Assessment." (n.d.) [http://128.138.129.98/education/faculty/edwardwiley/Docs/Wiley\\_A%20Practitioners%20Guide.pdf](http://128.138.129.98/education/faculty/edwardwiley/Docs/Wiley_A%20Practitioners%20Guide.pdf) (accessed October 20, 2010).
- Wiley, Edward W., Eleanor R. Spindler, and Amy N. Subert. Denver Pro Comp: An Outcomes Evaluation of Denver's Alternative Teacher Compensation System-2010 Report. <http://static.dpsk12.org/gemsnewprocompProCompOutcomesEvaluationApril2010final.pdf> (accessed 10/20/2010).
- Wilms, Wellford W. and Richard R. Chapleau. "The Illusion of Paying Teachers for Student Performance." In *Taking Sides: Clashing Views on Controversial Issues in Secondary Education*, edited by Dennis L. Evans, 191-194. Guilford: McGraw-Hill Companies, Inc., 2002.
- Wilson, Mark, ed. *Towards Coherence Between Classroom Assessment and Accountability: 103rd Yearbook of the National Society for the Study of Education, Part II*. Chicago: The University of Chicago Press, 2004.
- Winger, Tony. "Grading What Matters: No matter how lofty our espoused education goals, our grading practices reveal what we truly value." *Educational Leadership* 67, no. 3 (November, 2009): 73-75.
- Wood, David F., and Dan S. Green. "Managerial Experience with Merit Pay: A Survey of the Business Literature." In *Merit, Money and Teachers' Careers: Studies on Merit Pay and Career Ladders for Teachers*, edited by Henry C. Johnson, Jr., 121-136. Lanham: University Press of America, Inc., 1985.
- Wynn, Richard, Chris A. DeYoung, and Joanne Lindsay Wynn. *American Education*. 8th ed. New York: McGraw-Hill Book Company, 1977.

## Appendix A: Survey Results

*Online survey given to Sycamore River teachers from 5/27/2011-6/1/2011*

Survey results of 142 teachers out of approximately 530 teachers in Sycamore River.

Question 1: NCLB standardized tests do a good job assessing 21st Century Learning Skills.

Strongly Agree	2
Somewhat Agree	27
Somewhat Disagree	44
Strongly Disagree	69

Question 2: The impact of standardized tests on education in the last 20 years has been:

Very Beneficial	5
Somewhat Beneficial	26
Somewhat Detrimental	67
Very Detrimental	44

Question 3: Sycamore River's plan for implementing TAP teacher evaluations uses California Standardized Tests for half the teacher's evaluation. Consider how teaching for standardized tests will compare to the teaching of 21st century learning skills. In Sycamore River TAP classroom standardized testing be Equal with, Less than, Greater than 21st century learning skills:

Equal with	36
Less than	36
Greater than	70

Question 4: How much is a student's performance on standardized tests influenced by outside forces--either helped or hurt by things outside the realm of the teacher or school?

Not at all influenced	2
Somewhat influenced	43
Very Influenced	97

---

Question 5: Sycamore River should place more emphasis on teaching for Standardized Testing:

Strongly Agree	4
Somewhat Agree	13
Somewhat Disagree	45
Strongly Disagree	80

---

Question 6: Sycamore River should place more emphasis on teaching 21st Century Learning Skills:

Strongly Agree	62
Somewhat Agree	71
Somewhat Disagree	4
Strongly Disagree	5

---

Question 7: Does teaching for Standardized Tests and teaching 21st Century Learning Skills compliment or contradict each other. **(Unfortunately, I cannot count the results from this question, because I accidentally wrote strongly contradicts (twice), instead of writing strongly compliments)**

---

Question 8: What do you think are the findings of a majority of non-profit research experts REGARDING Multiple Choice Standardized Tests:

Are effective tools for assessing learning	4
Are not accurate, and narrow the curriculum	82
Are not perfect, but are necessary for comparisons	56

---

Question 9: If NCLB requirements were dissolved, would that change your opinion on these responses?

No	92
Yes	50

---

Any further comments (optional): Appendix B reports the optional comments by teachers.



## Appendix B: Survey Comments

**This is a collection of the comments provided by teachers in the comment box that was an optional part of the survey.**

For years I have been trying to narrow the gap in communication between test makers and teachers. How can there be such little clarity and dialogue within the education community at the opposite ends of the standard testing world? There is such disparity between test procedure makers and test givers. We have to sign an agreement NOT to even talk about that which we desperately need to have an open forum about!! The curriculum is almost entirely based upon a test which we can't modify or practically adjust to our teaching? No input, comments, or questions allowed? -- That is ridiculous!!

1.It's my understanding that at least part of the reason the TAP program exists is to improve test scores, so that [district name omitted] will be released from Program Improvement status. It stands to reason that teachers must teach to the test in order to do this and in order to receive merit pay.

2.I'm sure that studies show that the younger the child the more likely they are to score poorly if there was a problem at home (a pet died), if they had no breakfast, or if they are worried about the test itself. They are more easily distracted.

3. Our district has gone one step further than giving us the state test emphasis areas. They have given us pacing guides for language arts and math. We are now being told what page to teach on what day. Combination classes with two grade levels have little time to teach anything else but math and language skills to keep on track. Math takes about 1 1/2 hours a day alone.

4. We buy thousands of dollars of books and other materials when a school goes into PMI. New materials are not the answer. Find the teachers who don't do well and help them. Give principals more time to help them.

5. I bet all studies show that the best way to improve student learning is smaller class sizes.

The teaching of both 21st Century skills as well as state standards is essential for student learning. Teachers need to be responsible for the education they are providing students. Standardized tests and core standards help to create a consistent curriculum and ensure that all students are learning essential skills. A teacher must incorporate 21st Century skills in his/her curriculum so that students are also prepared to compete in a global economy. In order for all of this to happen, we must change the way we educate students and focus on keeping good teachers in the classroom and getting bad teachers out.

Once teachers lose the personal touch, teaching will be doomed. It is the connection between teacher and student and the passion the teacher has for teaching and the subject that will be the most effective on students. Teaching to facts and tests hurts our students and our nation.

Standardized testing is making our students LESS competitive in the global marketplace. We are not adequately preparing our students for the real 21st century world.

I hate standardized tests!

For mathematics, there is so much building that occurs. There is overlap of what is tested and what are 21st Century Learning Skills in application. My biggest problem is with what is being expected at the Pre-Algebra level, with its vastness of curriculum covered. If the CA Standards were refocused and revamped and then addressed as we currently are trying to meet them as is, I believe more success could be enjoyed by all, and the students would be the biggest winners.

Which is more important, test scores or students being able to think out side the box to solve a problem? As long as tests scores drive our schools, independent and group creativity in problem solving will slowly fall to way-side. And that will be very sad.

I am a kindergarten teacher and I do not give the standardized tests but I do know that these tests are helpful in tracking the students who are above grade level but they can be detrimental for the students who are achieving below grade level because it's just another opportunity for them to fail again. The tests can also be devastating for the child who is at grade level because they are testing material above their grade level and they get deflated when they can not do it. As far as the twenty first century skills, I find it very interesting that they want to implement things without the proper resources, for example, no computer lab techs at school sites, bilingual education without bilingual teachers. These are just my thoughts, I hope they were helpful.

I don't believe TAP will limit teachers teaching 21st century skills. I hope TAP will give teachers new insight on teaching strategies and help teachers engage students and motivate students to want to learn. Becoming a better teacher should be the focus of TAP, not teaching to a test. If teachers are doing a better job engaging and motivating students they should learn more and do better on valid assessments.

Testing dehumanizes our students and takes time away from aha moments.

No Child Left Behind has ruined the education of our students.

I don't mind the testing. I just mind the consequences. It amazes me that the future of my school could be hanging on the whims of 7 to 12 year olds.

This survey was incredibly narrow. There was no response for "No Opinion." Standardized testing is a thin slice of a student's academic knowledge. It is an easy thing to do and gives us one view of what a student knows and as such has some value as long as its limitations are acknowledged.

I have taught 14 years @ [district name omitted], but at the same sight. So my answers reflect my experience at this one sight only.

This survey is not going to be an accurate reflection of our thoughts! There is no neutral position on any question.

It's hard to be it all and teach it all in a limited day, isn't it? I do still feel it is possible to utilize teaching of technology and 21st century skills to help build student connection to curriculum and help build performance on testing...but that will be challenging and not easily implemented by everyone - especially considering many of our teachers do not have a strong technology background to bring to the teaching of these skills. It may be difficult for them to determine methods to implement technology in the classroom to help build standardized testing performance. I do not like that the curriculum is being narrowed by the emphasis on testing, but the fact is that's the world we presently live in and the hoops we must jump through. If we don't jump we get punished, so jump we must.

Setting standards, some curriculum guidelines, and common assessments naturally have some merit to insure all teachers are getting their students prepared for the foundation on which the next grade level material must be built upon. The testing has been over kill. Teaching to the tests, preparing for the tests, and administering the tests has taken a great deal of time out of real learning. I have seen students with zero English randomly bubble in tests over the years and score "basic" or near basic. I have seen students add up all the side of a perimeter, get the answer wrong, choose something close and get the answer right. I am embarrassed when a student scores higher than their ability level, which often happens with multiple choice questions. I know I did not help the student at all, but high scores are always suspect. Granting rewards for higher test scores will only promote "help" for the desperate ones to look good on paper. The real assessment is a teacher's judgement. Teachers KNOW if a student is high, medium or low in any given subject. The number game is not very valuable. Bubble in tests do not promote deep critical thinking, which is lacking in a lot of students today.

Teaching "to the test" should be outlawed. It has caused real learning to decrease because of the emphasis on having children memorize minutia at its worst and learning test taking strategies at the best. what a waste of time and money and energy of everyone involved.

As long as monies and teacher effectiveness( don't tell me it's not) is based on standardized test scores, we will need to work with this necessary evil.

Standardized CST testing took three plus weeks out of our school year, and effectively let the air out of the tires for the motivation we had built to continue learning the rest of the year.

I taught this year toward the outcomes on the test. In the past I have spent more time on critical thinking, logic and analyzing information.

The survey questions seem to imply that 21st Century Learning Skills are a logical counterpoint or alternative to NCLB dictates. I tend to think of the 21st Century Skills mandate, as promoted by [district name omitted], to be another pre-packaged concept

lacking in nuance and the necessary differentiation needed to truly address the educational needs of our students. For example, I think there is a tendency to push technology for technology's sake, and not because it is shown to enhance learning or increase a desired skill. How many of us have sat through a Power Point presentation which uses the technology correctly but fails to interest or advance an idea? Power Point is a useful tool, but overused and rarely used to good effect. As a math teacher, I know that it's important for students to see the PROCESS of mathematical computations being performed step-by-step. One of my advanced high school math students reported to me that she had a hard time learning from a middle school teacher who relies on a Smart Board to present his lessons. She said the problems were beautifully presented on the Smart Board, correctly worked out and complete, but she had trouble following the process. As a math teacher, I use very little technology in the lower level courses because the greater risk is that students become over reliant on technology--specifically calculators--to the detriment of their critical thinking and problem solving skills. The need for those skills hasn't changed much in the 21st Century.

Instead of multiple choice, like this survey, a much more accurate assessment is free response and short answer. Standardized tests are not a problem. It is simply the format of the standardized test (multiple choice) that results in an inaccurate assessment of student achievement. The current format also kills the opportunity for creative and critical thinking.

The primary purpose in having a credentialed teacher in the classroom is to insure that the teacher's academic skills are being directly transmitted to the students at their grade level. To follow up and accurately measure whether this is being done, districts must develop measurement tools, aka 'tests', to find out how effective the teaching and learning is. This is why I do believe in standardized testing.

Teaching basic skills is as important as teaching 21st Century Learning Skills!

The hiring and firing of principals, implementation of all curriculum, all supplemental instruction, and the writing of all california education adoptions are based upon the premise, either real or false, that they will raise test scores. Nothing happens in our school district that is not based on passing those tests, at least at my school which is in danger or becoming stage 3 non compliant in it's scores

Until tests are proctored, cheating can occur. Tests are invalid measure of learning.

[district name omitted], with its Target Quizzes and control of subject pacing guides, is strongly teaching to the test. This attitude is taking all the fun out of teaching and all the joy out of learning. Students moan every time we do those quizzes or practice tests and I can't blame them. It is getting harder and harder to make learning fun because there is so little flexible time. I honestly don't see how project-based learning can function well in [district name omitted] with all the time budgeted to the TEST.

---

### Appendix C: Interview Details and Format

From May 16-June 4, 2011, a total of 23 teachers and 2 administrators were interviewed--all from Sycamore River, except for one neighboring district high school teacher. The teachers from Sycamore River represented nearly all grade levels and nearly all subjects. Each interview lasted anywhere from four minutes to twenty four minutes. The total length of time for all the interviews is over four hours. The teacher from outside the district expressed the same general concerns as Sycamore River teachers expressed. These interviews are documented on video camera and recorded onto four DVDs in the possession of Shawn Greenelsh and Dr. Joel Orth and Dr. Roberta Herter. The names of these teachers and or administrators are not allowed to be used outside the confines of California Polytechnic State University, San Luis Obispo.

Below is the information and questions that led each interview:

Background Information: In the last two decades, policymakers have implemented measures that elevate the status of standardized testing to a high-stakes level, like Program Improvement measures in No Child Left Behind. Another current trend is teacher merit pay systems that rely on standardized test results to evaluate teachers.

---

Please tell me what you teach and how long have you been teaching.

---

1. What type of learning outcomes are standardized tests good at assessing?  
What types of learning outcomes are not assessed by standardized tests?

---

2. How much is a student's performance on standardized tests influenced by outside forces--either helped or hurt by things outside the realm of the teacher or school? Explain.

---

3. How have standardized tests and the ramifications from the results of those tests shaped education in the last twenty years?

---

4. *Differing Beliefs:*

If you think standardized testing and merit pay are generally the best tools we have to measure student and teacher performance, then why do you think some teachers and administrators advocate project-based learning and authentic assessment?

If you think standardized testing does a poor job measuring student and teacher performance, then why do you think some teachers and administrators advocate it?

---

5. What do you think the research experts say about:  
a. Standardized Testing  
b. Teacher merit pay systems that use  
standardized test-results to evaluate a teacher?

---

## **Appendix D: A Brief Description of TAP<sup>214</sup>**

### **What are the key elements of TAP?**

1. Multiple Career Paths: Each elementary school would have one fully-released master teacher and two mentor teachers, while each middle school would have two master teachers and three mentor teachers. Master teachers work an extra 20 days, and mentor teachers work an extra 10 days.
2. On-going professional development: The school schedule is modified to provide about an hour each week for cluster meetings at each grade level or department level. This is a form of professional learning community where teachers look at student data, plans lessons, and learns new instructional strategies under the direction of the master and mentor teachers. Literacy Support Teachers would provide coverage of the classrooms on a weekly basis.
3. Instructionally Focused Accountability: Teachers are observed at least four times a year by the administrator, master or mentor teacher using research-based rubrics for several dimensions of instructional quality. Observers are trained and certified on these rubrics. The principal is still responsible for the actual teacher evaluation.
4. Performance-based Compensation: Teachers in a TAP school have the opportunity to earn bonuses averaging about \$3,000 per teacher at each site. The bonuses are based on rubric scores, students' achievement gains and school achievement growth. In TAP schools across the nation, on average over 90% of the teachers receive a bonus.

The TAP rubric will not be used to fire teachers!! The purpose of TAP is to provide teachers and principals with coaching and feedback. We already have a PAR system in place for teachers needing assistance; the TAP system will not replace this. We are not trying to get rid of the current seniority system or the current tenure system with TAP- that is not part of our intention at all.

Put simply, our district has seen a 20% reduction in state funding over the past three years. Unfortunately for us, these economic times collided with an increasing level of accountability that is unprecedented. We are in district-wide Program Improvement sanctions, and have seven individual schools that are in Program Improvement. While money cannot solve all of our challenges, adding instructional coaches, providing structured release time, and accessing some of the best and most creative teachers across the nation (through the TAP online library of strategies and videotaped demonstration lessons) can certainly help.

---

<sup>214</sup> Description from the Superintendent of Sycamore River in his full address to the district staff, Superintendent News, October 1, 2010. This excerpt pulled from the district website [the district website has been omitted for anonymity reasons] (accessed June 14, 2011).